Radiative Processes in Astrophysical Gases: From the Intergalactic and Interstellar Medium to Exoplanetary Atmospheres

Thesis by
Antonija Oklopčić

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended May 9, 2017

Antonija Oklopčić
ORCID: 0000-0002-9584-6476

Mojim roditeljima,
bez čije ljubavi i podrške ne bih bila to što jesam.

ACKNOWLEDGEMENTS

Time that I spent as a graduate student at Caltech has been an amazing period in my life and I am grateful to everyone who has been a part of it. I would like to begin by thanking everyone who has helped me get here in the first place, especially my undergraduate advisor Vernesa Smolčić. I gratefully acknowledge the support from the Fulbright International Science & Technology Fellowship that I have received during the first three years of graduate school.

Above all, I would like to thank my advisor, Chris Hirata, for his immense kindness and support throughout the past six years, and especially for giving me the freedom to pursue my own scientific interests, wherever they may take me. I greatly enjoyed working with Chris and learning from him—I am very honored and proud to be his student.

I owe gratitude to Phil Hopkins for introducing me to the world of hydrodynamic simulations. It has been great fun and privilege to work with him and the rest of the FIRE collaboration. I am grateful to Phil and his large group of students and postdocs for many interesting discussions, group meetings, and informal workshops—I have learned so much from them.

I thank the members of my thesis defense committee: Lynne Hillenbrand, for always being supportive of students and committed to their well-being, especially in her role as the option representative; Tony Readhead, for his generosity and kindness, as well as for introducing me to the wonders of radiative processes in Ay121; Dimitri Mawet, for his interest and enthusiasm.

I am grateful to Kevin Heng, who was the first to encourage Chris and me to start thinking about problems related to exoplanet atmospheres. Ever since, Kevin's input and advice have been of immense value for both my research and for choosing my career path.

My graduate school experience would not have been the same without the amazing friends that I have acquired along the way. I am extremely grateful to Gwen Rudie, my friend and mentor from day one. Gwen has been an unlimited source of wisdom, compassion, and support in my life for the past six years and a real role model. Ben Montet, Allison Strom, Trevor David, and Yi Cao – you guys made taking classes, doing homework, and studying for the qual more fun than I ever thought it could be possible. I will forever cherish the memories of the adventures we had together—

camping in the Grand Canyon, wine tasting in Napa, enjoying the fireworks at a Star Trek-themed Dodgers game, the first-year Christmas challenge(s), and many more. Thank you for being such an important part of my life! Abhilash Mishra and Teja Venumadhav, my academic brothers, thank you for all the fun we had together in Pasadena, Columbus, and everywhere in between. And finally, many thanks to all graduate students in Caltech Astronomy and TAPIR, past and present, for countless lunches at Chandler, Thursday donuts and Friday beers, Halloween parties, grad formals, softball games, and ski trips.

To my favorite fellow grad student—my husband Mislav—I cannot express how grateful I am to you for everything in our lives. I know I would not be where I am today if it hadn't been for your love, support, and encouragement along the way. I love sharing my life with you and I am looking forward to the next big step ahead of us—Cambridge (Massachusetts, not England), here we come!

Last but not least, I would like to thank my family—my parents and my sister—for their unconditional love and support in everything I do. Thank you for encouraging me to follow my dreams and for teaching me to value knowledge, education, and hard work. And to never lose hope.

ABSTRACT

This thesis presents theoretical investigations in three areas of astrophysics, all related to radiative processes and interactions between stellar radiation and gaseous media in the Universe, ranging from the intergalactic and interstellar medium to planetary atmospheres.

Part I of the thesis consists of two independent investigations in which we study the effects of stellar feedback in high-redshift environments. The topic of Chapter 2 is the intergalactic medium (IGM) in the epoch just after the formation of the first stars in the Universe, but before the cosmic reionization was completed. This epoch is of great interest for the ongoing and future experiments aimed at observing the neutral IGM via the redshifted 21 cm line of hydrogen. We study the effects of resonant scattering of Lyman-α photons produced by early stars on the structure of temperature fluctuations in the IGM. In Chapter 3, we use cosmological hydrodynamic simulations of galaxy evolution to study the effects of stellar feedback on the clumpy structure of star-forming galaxies at $z \sim 2$. Observations of high-redshift galaxies show that their morphology is often dominated by a few giant clumps of intense star formation, but the nature and the importance of these clumps for the evolution of their host galaxies are uncertain. We present a detailed analysis of the properties of giant clumps in a high-redshift simulated galaxy from the FIRE project.

Part II of the thesis is devoted to the effects of Raman scattering of stellar radiation in the atmospheres of extrasolar planets. Spectral signatures of Raman scattering imprinted in the geometric albedo spectrum of a gaseous planet carry information about the properties of the planet's atmosphere—its composition, temperature, and the radiation-penetration depth. In Chapter 5, we present the results of radiative transfer calculations including the treatment of Raman scattering for different types of planetary atmospheres and analyze the feasibility of detecting the spectral signatures of Raman scattering in nearby exoplanets. The structure and the intensity of Raman spectral features depends on both the atmospheric properties and the shape of the stellar spectrum irradiating the atmosphere. In Chapter 6, we analyze the diversity of Raman features in the geometric albedo spectra of planets hosted by different types of stars.

PUBLISHED CONTENT AND CONTRIBUTIONS

Oklopčić, A., Hopkins, P. F., Feldmann, R., et al., 2017,
"Giant clumps in the FIRE simulations: a case study of a massive high-redshift galaxy"; Monthly Notices of the Royal Astronomical Society, 465, 952
DOI:10.1093/mnras/stw2754
A.O. co-developed the idea for the project, performed the analysis, and wrote the paper.

Oklopčić, A., Hirata, C. M., & Heng, K., 2016,
"Raman Scattering by Molecular Hydrogen and Nitrogen in Exoplanetary Atmospheres"; The Astrophysical Journal, 832, 30
DOI:10.3847/0004-637X/832/1/30
A.O. developed the idea for this project, performed the analysis, and wrote the paper.

Oklopčić, A., & Hirata, C. M., 2013,
"Lyα Heating of Inhomogeneous High-redshift Intergalactic Medium"; The Astrophysical Journal, 779, 146
DOI:10.1088/0004-637X/779/2/146
A.O. performed the analysis for this project and wrote the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Part I

Stellar Feedback Effects in the Intergalactic and Interstellar Medium at High Redshift

Chapter 1

INTRODUCTION TO PART I

After the hydrogen recombination, which occurred approximately 400,000 years after the Big Bang, the Universe consisted of mostly neutral gas of primordial composition – roughly 75% hydrogen and 25% helium, with trace amounts of lithium (Cyburt et al. 2016). The ground state of the most abundant element in the Universe – the hydrogen atom – consists of two hyperfine levels separated in energy by $\Delta E \sim 5.87\ \mu$eV, which corresponds to a photon wavelength of ~ 21.1 cm. Observation of the redshifted 21 cm signal of neutral hydrogen is regarded as the most promising way to observe the early Universe before the formation of the first sources of radiation (i.e. the first stars and galaxies). Reviews of the 21 cm physics and the efforts to observe the 21 cm radiation form high redshift are given in Furlanetto et al. (2006), Morales & Wyithe (2010), Pritchard & Loeb (2012), and Barkana (2016). A brief overview presented in this chapter is mostly based on these papers.

The nature and the intensity of the expected 21 cm signal depends on the difference between the temperature of the cosmic microwave background (CMB) radiation and the excitation temperature of the hydrogen hyperfine transition, which is usually referred to as the spin temperature (T_S) and is determined by the relative population of the upper (n_u) and the lower (n_l) level of the hydrogen ground state:

$$\frac{n_u}{n_l} = 3 \exp\left(-\frac{\Delta E}{k T_S}\right). \tag{1.1}$$

If the spin temperature is lower (greater) than the temperature of the CMB, the 21 cm signal can be observed in absorption (emission) against the background radiation. The 21 cm signal cannot be observed if the two temperatures are equal. Therefore, in order to predict the expected 21 cm signal from the early Universe, it is important to have a good understanding of how the spin temperature evolves with time.

In a dense environment of the early Universe, collisions between atoms happen at a high enough rate to keep the spin temperature coupled to the gas kinetic temperature (T_K). In the first $\sim$ 10 Myr after the recombination, the residual amount of free electrons in the IGM is sufficient to keep the kinetic temperature coupled to the CMB temperature through Compton scattering of the CMB photons and electrons.

In other words, in the redshift range between the recombination ($z \sim 1,100$) and $z \sim 150 - 200$, the spin temperature is equal to the temperature of the CMB, thus making this period in the evolution of the Universe unobservable via the 21 cm radiation.

As the Universe expands and cools, at $z \sim 150 - 200$ the ionization fraction of the IGM drops and so does the rate of Compton scattering of the CMB photons. Since the IGM can no longer stay coupled to the CMB, it adiabatically cools[1] according to $T_K \propto (1 + z)^2$. The IGM density is still high enough to couple T_S to T_K through collisions, causing $T_S \propto (1 + z)^2$. The temperature of the CMB photons (T_γ), on the other hand, falls at a slower rate, with the redshift dependence $T_\gamma \propto (1 + z)$. Therefore, at this point in its evolution, the Universe becomes observable via the redshifted 21 cm radiation.

Around $z = 30$, the density of the IGM drops to a level at which collisions between atoms become so rare that the spin temperature cannot stay coupled to the gas kinetic temperature any longer. If the CMB is the only source of radiation, the spin temperature again becomes equal to the CMB temperature, and the 21 cm signal is lost once again. However, if there are other sources of radiation, the story becomes entirely different. The spin temperature can be coupled to the radiation field temperature via the scattering of Lyman-α (Lyα) photons, which is known as the Wouthuysen-Field effect. Lyα scattering can change the relative population of the hyperfine levels if, upon scattering, the hydrogen atom does not return to the same ground-state level as it originated from. Small initial density perturbations in the IGM grew over time and around $z \sim 20 - 40$ the first collapsed objects in the Universe started to form. As the first stars and galaxies turn on, their radiation changes the spin temperature of the IGM (through the Wouthuysen-Field coupling), its kinetic temperature as the gas heats up, and eventually its ionization state as the IGM becomes (re)ionized.

Because of this complicated interplay between the gas temperature, the spin temperature and the temperature of the radiation field, the expected 21 cm signal from this epoch depends on many details of the cosmic structure formation and its influence on the evolution of the IGM. The effects of this early stellar feedback, taking place

[1] For an adiabatically expanding ideal gas, the pressure and volume are related through PV^γ = constant, where γ is the adiabatic index equal to $5/3$ for the monoatomic IGM. Using the ideal gas law and $V \propto (1 + z)^{-3}$, it is straightforward to get $T \propto (1 + z)^2$. The condition of constant entropy for the photon gas is satisfied when $PV^{4/3}$ = constant, which has a similar form to the ideal gas expression, but different origin, since the ideal gas approximation does not apply to photons. The following temperature-redshift dependence for the CMB photons is $T \propto 1 + z$.

even before the formation of the first galaxies in the Universe, are discussed in the next section, which provides background for the work described in Chapter 2.

1.1 Cosmic Dawn

The epoch in the evolution of the Universe in which the first luminous objects (the first stars and galaxies) formed is often referred to as the *cosmic dawn*. In this section, I give a brief overview of our current understanding of this epoch, which is very limited due to the lack of direct observations; for a comprehensive review of this topic, see Loeb & Furlanetto (2013).

Primordial density fluctuations, imprinted as temperature fluctuations in the CMB, grew over time due to gravity. When dark matter forms a collapsed object (a halo), the co-located baryons fall into the halo, experience shocks and get heated to the virial temperature of the halo. The gas will not undergo further gravitational collapse if it is supported by thermal pressure. Therefore, in order to fragment and form stars, the gas first needs to cool. In the local Universe, cooling primarily happens through radiative transitions of metals. The primordial gas, however, does not posses any metals (except trace amounts of lithium), so it can only cool via radiative transitions of atomic or molecular hydrogen.

Molecular cooling begins when objects of $\sim 10^6$ $M_\odot$, the so-called minihalos, start to collapse. The gas in minihalos loses its thermal energy via radiative cooling, fragments and forms the first generation of stars in the Universe. The Lyman-Werner radiation (i.e. photons with energy in the range from 11.2 to 13.6 eV), emitted by these stars, can destroy H_2 molecules in nearby halos and thus postpone further cooling and star formation until the formation of much larger halos, in which gas can cool via radiative transitions of atomic hydrogen. Halos of mass on the order of 10^8 $M_\odot$have the virial temperature of about 10^4 K, which is sufficient for hydrogen atoms to get collisionally excited and radiate.

The early generation of stars, especially those capable of producing high-energy (UV and X-ray) radiation, such as the massive stars that form X-ray binaries, had a profound influence on the evolution of the IGM. X-rays can heat the IGM by photoionizing the atoms and producing energetic free electrons. These electrons dissipate their kinetic energy by transferring it to the surrounding gas and thus heating it up. Scattering of UV photons, in particular Lyα photons, can also cause heating of the gas due to the atomic recoil during the scattering process. Atoms get a 'kick' for every scattering event they experience. The energy transferred to the

gas in a single scattering is small, but the number of scatterings per atom is large, so the cumulative effects can be significant.

Even though it is generally believed that X-ray heating dominates over Lyα heating on large scales because X-ray photons have a greater mean free path and can impact the gas farther away from the source of radiation, Lyα heating may play an important role in changing the temperature structure of the IGM on smaller scales. This topic is further explored in Chapter 2, in which I present an analysis of how the presence of density, temperature, and velocity perturbations in the IGM affects the efficiency of Lyα heating on small scales.

1.2 Cosmic Noon

After the formation of the first stars and galaxies, the structure in the Universe continued to grow hierarchically – larger and lager halos formed via mergers of smaller objects and through accretion of fresh material supplied via the cosmic web filaments. It is believed that the photons escaping the early galaxies caused reionization of the Universe, which finished around $z \sim 6$. Many important aspects of cosmic reionization are still unknown and are major subjects in modern cosmological research.

Large sky surveys, conducted over the past decade with numerous telescopes and over a wide range of wavelengths, provide the following picture of the cosmic star formation history (for a review see Madau & Dickinson 2014). The density of cosmic star formation had a rising phase at $3 \lesssim z \lesssim 8$, reached its peak around $z \sim 1.5 - 3$ (about ~ 10 Gyr ago), and has been experiencing a decline ever since. The peak epoch, which is often referred to as the *cosmic noon*, is a particularly interesting period in the evolution of the Universe – stars were forming at about an order of magnitude higher rate than today and the star-forming galaxies of the period had a different morphology than star-forming galaxies in the local Universe (Conselice 2014).

Massive star-forming galaxies at high redshift ($z \sim 2$) look much more irregular compared to present-day galaxies of similar mass, which typically have well-defined star-forming disks overlaid with overdense regions in the form of spiral arms. The morphology of high redshift galaxies if often clumpy and dominated by a few bright star-forming regions scattered across the underlying continuum. Since this unusual morphology of high-redshift galaxies was first noticed in the mid-1990s (e.g. Cowie et al. 1995; van den Bergh et al. 1996), the properties of these star-

forming clumps and their relevance for galaxy evolution have been investigated in numerous theoretical and observational studies (for more details see Section 3.1 and references therein).

Using hydrodynamic simulations of galaxy evolution seems like the most straightforward way of studying the nature of giant clumps. However, the results heavily depend on the details of sub-grid prescriptions used to describe stellar feedback in simulations. For example, in simulations that implement stellar feedback as thermal energy from supernova explosions, giant clumps form stars very efficiently and stay bound as dense clusters for more than $\sim 10^8$ yr (e.g. Ceverino et al. 2010). During that period, the clumps spiral inwards to the center of their host galaxy due to dynamical friction. This radial migration of large star clusters formed within galaxies themselves can be viewed as an alternative mechanism for the formation of galactic bulges, which may have important consequences for our understanding of how galaxies evolve through cosmic time. On the other hand, simulations that have more intense stellar feedback prescriptions, especially in the form of strong radiative pressure feedback during the early stages of stellar life, show that clumps get disrupted on a relatively short time-scale ($\sim 10^7$ yr), which does not allow for significant radial migration (Genel et al. 2012; Hopkins et al. 2012). Giant clumps in these simulations do not have a major effect on bulge formation. Therefore, in order to determine whether or not clumps play an important role in galaxy evolution, it is crucial to model the effects of stellar feedback in hydrodynamic simulations as realistically as possible.

The importance of feedback in galaxy formation and evolution has been known for several decades (see review by Somerville & Davé 2015). Basic models of structure formation within the cold dark matter framework suffer from an overcooling problem – they produce too much stars compared to observations. Feedback is believed to explain why the baryon fraction within galaxies is smaller than the average cosmic value and why the global star formation efficiency is low. By stellar feedback we usually refer to various physical processes involving (massive) stars and supernovae that transfer copious amounts of energy and momentum to the interstellar medium (ISM). Stellar feedback regulates star formation within a galaxy via its two main effects: (1) heating and expelling the gas from within the galaxy, and (2) preventing the surrounding gas from cooling and accreting onto the galaxy.

Modeling stellar feedback in numerical hydrodynamic simulations is complicated because it involves predicting large-scale effects of processes that happen on spatial

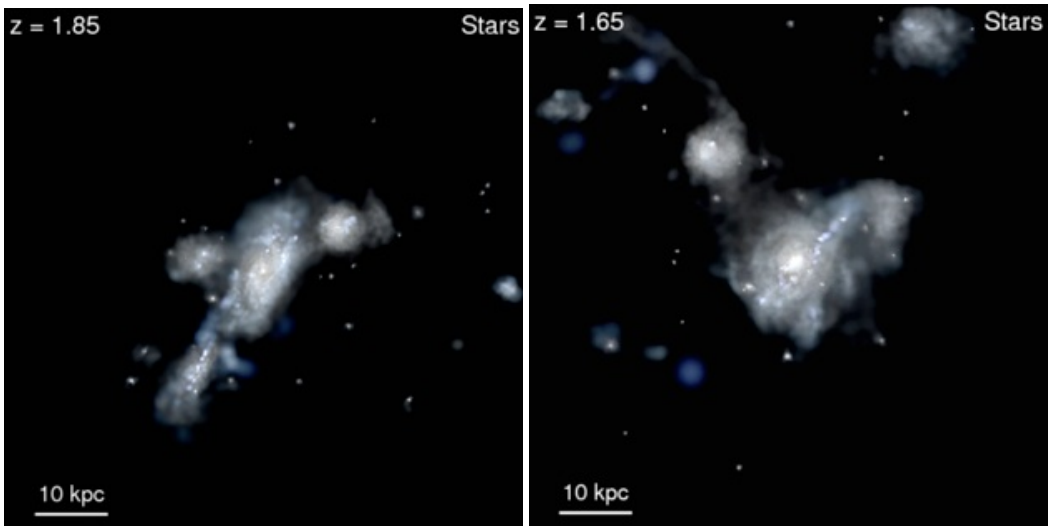

Figure 1.1: Mock three-color images (u/g/r bands) of a galaxy from the FIRE project (simulation `m12v`) undergoing mergers at high redshift (around the epoch of cosmic noon). Images made by Philip F. Hopkins and available at http://fire.northwestern.edu/visualizations/.

and time scales much smaller than those resolved in simulations of entire galaxies or even larger volumes (i.e. cosmological simulations). Early attempts included just thermal energy from supernovae and they failed to produce galaxies whose properties matched those of the observed galaxies (e.g. Katz et al. 1996). However, simulations have been improving over the years.

The FIRE project (*Feedback In Realistic Environments*, Hopkins et al. 2014; Hopkins et al. 2017) is dedicated to exploring the effects of many different channels of stellar feedback on galaxy evolution. The project consists of a suite of high-resolution cosmological zoom-in simulations incorporating explicit treatments of stellar feedback and the multi-phase ISM physics. These simulations reproduce many observed scaling relations, as well as the physical and morphological properties of real galaxies, without tuning the simulation parameters "by hand" to match observations. The synergy of high resolution and the realistic ISM physics, makes the FIRE simulations an ideal testing ground for studying the properties and evolution of giant star forming clumps in high-redshift galaxies, which is the topic of Chapter 3.

References

Barkana, R., 2016, PhR, 645, 1

Ceverino, D., Dekel, A., & Bournaud, F., 2010, MNRAS, 404, 2151

Conselice, C. J., 2014, ARA&A, 52, 291

Cowie, L. L., Hu, E. M., & Songaila, A., 1995, AJ, 110, 1576

Cyburt, R. H., Fields, B. D., Olive, K. A., & Yeh, T.-H., 2016, Reviews of Modern Physics, 88.1, 015004, 015004

Furlanetto, S. R., Oh, S. P., & Briggs, F. H., 2006, PhR, 433, 181

Genel, S., Naab, T., Genzel, R., et al., 2012, ApJ, 745, 11, 11

Hopkins, P. F., Kereš, D., Murray, N., Quataert, E., & Hernquist, L., 2012, MNRAS, 427, 968

Hopkins, P. F., Kereš, D., Oñorbe, J., et al., 2014, MNRAS, 445, 581

Hopkins, P. F., Wetzel, A., Keres, D., et al., 2017, ArXiv e-prints

Katz, N., Weinberg, D. H., & Hernquist, L., 1996, ApJS, 105, 19

Loeb, A., & Furlanetto, S. R. 2013, The First Galaxies in the Universe

Madau, P., & Dickinson, M., 2014, ARA&A, 52, 415

Morales, M. F., & Wyithe, J. S. B., 2010, ARA&A, 48, 127

Pritchard, J. R., & Loeb, A., 2012, Reports on Progress in Physics, 75.8, 086901, 086901

Somerville, R. S., & Davé, R., 2015, ARA&A, 53, 51

van den Bergh, S., Abraham, R. G., Ellis, R. S., et al., 1996, AJ, 112, 359

Chapter 2

LYMAN-α HEATING OF INHOMOGENEOUS HIGH-REDSHIFT INTERGALACTIC MEDIUM

ABSTRACT

The intergalactic medium (IGM) prior to the epoch of reionization consists mostly of neutral hydrogen gas. Lyman-α (Lyα) photons produced by early stars resonantly scatter off hydrogen atoms, causing energy exchange between the radiation field and the gas. This interaction results in moderate heating of the gas due to the recoil of the atoms upon scattering, which is of great interest for future studies of the pre-reionization IGM in the HI 21 cm line. We investigate the effect of this Lyα heating in the IGM with linear density, temperature and velocity perturbations. Perturbations smaller than the diffusion length of photons could be damped due to heat conduction by Lyα photons. The scale at which damping occurs and the strength of this effect depend on various properties of the gas, the flux of Lyα photons and the way in which photon frequencies are redistributed upon scattering. To find the relevant length scale and the extent to which Lyα heating affects perturbations, we calculate the gas heating rates by numerically solving linearized Boltzmann equations in which scattering is treated by the Fokker-Planck approximation. We find that (i) perturbations add a small correction to the gas heating rate, and (ii) the damping of temperature perturbations occurs at scales with comoving wavenumber $k \gtrsim 10^4$ Mpc^{-1}, which are much smaller than the Jeans scale and thus unlikely to substantially affect the observed 21 cm signal.

This chapter has been published as Oklopčić, A. & Hirata, C. M. 2013, ApJ, 779, 146.

2.1 Introduction

After recombination of the primordial plasma at redshift $z \approx 1100$ and before the epoch of reionization, the baryonic content of the Universe was predominantly in the form of neutral hydrogen. For this reason, a promising way of probing this period in the evolution of the Universe is through the observations of the redshifted 21 cm line of neutral hydrogen, created in the spin-flip transition between the two hyperfine levels of the hydrogen ground state (for reviews of the 21 cm physics, its use in cosmology, and the foreground and calibration challenges see, for example, Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012). There are several experiments that are currently in operation, or planned for the near future, for which a primary objective is observing the redshifted 21 cm signal, such as the Low Frequency Array[1] (LOFAR; van Haarlem et al. 2013), the Murchison Widefield Array[2] (MWA; Lonsdale et al. 2009), the Precision Array to Probe EoR[3] (PAPER; Parsons et al. 2012), the Long Wavelength Array[4] (LWA) and the Square Kilometer Array[5] (SKA; Rawlings & Schilizzi 2011). Several pathfinder observations have placed upper limits on the 21 cm perturbation signal at $z = 7.7$ (Parsons et al. 2014), $z = 8.6$ (Paciga et al. 2013) and $z = 9.5$ (Dillon et al. 2014), and measurements of the global spectrum have placed lower limits on the duration of the neutral-to-ionized transition (Bowman & Rogers 2010).

The 21 cm signal from high-redshift intergalactic medium (IGM) is sensitive to the conditions of the gas, such as its density, ionization fraction and spin temperature. The last can be coupled to the gas kinetic temperature through collisions (in environments of sufficiently high density) or via the Wouthuysen-Field effect (Wouthuysen 1952; Field 1958) in the presence of Lyman-α (Lyα) photons. Hence, understanding thermal properties of high-redshift IGM is crucial for predicting and interpreting the observed 21 cm signal.

Before the formation of the first sources of radiation, the first stars and galaxies, primordial gas was adiabatically cooling with the expansion of the Universe–its temperature decreasing with redshift z as $(1 + z)^2$. The onset of luminous structures dramatically changed that evolutionary track and the Universe eventually became reheated and reionized. In a complete model of reionization, several different

[1] www.lofar.org
[2] www.mwatelescope.org
[3] eor.berkeley.edu
[4] www.phys.unm.edu/ lwa
[5] www.skatelescope.org

mechanisms can affect the temperature and the ionization state of the IGM, such as heating by X-ray and UV photons as well as by shocks created in the gravitational collapse of matter. In this work, we focus on the microphysics of the IGM heating through its interaction with the UV photons.

Non-ionizing photons emitted by stars can freely travel through mostly neutral high-redshift IGM until they are redshifted by the expansion of the Universe into a resonant frequency of one of the atomic species present in the IGM, at which point they resonantly scatter with atoms. This scattering can occur far away from sources of radiation. Since hydrogen is the most abundant element in the Universe and is mostly in its ground state at high redshifts prior to reionization, the most significant resonance is the Lyα transition between the ground state and the first excited state of hydrogen ($\lambda_\alpha = 1216$ Å, $\nu_\alpha = 2.47 \times 10^{15}$ Hz). There are two types of Lyα photons that need to be taken into account: photons emitted with frequencies between Lyα and Lyβ can redshift directly into the Lyα resonance, whereas photons of higher energies (between Lyγ and the Lyman limit) can fall into one of the higher Lyman resonances, from which they can cascade into Lyα (see e.g. Pritchard & Furlanetto 2006; Hirata 2006). To distinguish these two types of photons, we call the first kind the *continuum photons* and the second kind the *injected photons*.

The resonant scattering of Lyα photons with hydrogen atoms causes transfer of energy from the radiation field to the gas due to atomic recoil, causing a change in the kinetic temperature of the gas (Madau et al. 1997; see also Chen & Miralda-Escudé 2004; Furlanetto & Pritchard 2006; Meiksin 2006; Ciardi & Salvaterra 2007). This energy exchange leads to a drift of photons from higher to lower frequencies. Another contribution to frequency drift, one that is present regardless of scattering, comes from the Hubble expansion of the Universe. Scattering also causes a diffusion of photons in frequency space due to a Maxwellian distribution of atomic velocities in the gas. Photons on the red side of the Lyα line center mostly scatter with atoms moving toward them, because of the Doppler shift the frequency of these photons is higher in the frame of the atom; in other words, it is closer to the line center and the resonant frequency. The opposite occurs for the photons on the blue side of the line–they preferentially scatter off atoms moving away from them. Frequent scattering between atoms and photons brings them closer to statistical equilibrium, reducing the average energy exchange per scattering (Chen & Miralda-Escudé 2004). Fluctuations in the temperature and density of the gas, as well as gradients in its velocity, can change the Lyα scattering rates (Higgins &

Meiksin 2009) and thereby affect the heating of the gas by Lyα photons. Therefore, it is of great interest to understand how the theory of IGM heating by Lyα photons extends to the case of a realistic, inhomogeneous Universe—whether the heating rate is just slightly modified by the perturbations, or whether effects such as thermal conduction can become important.

In this study we investigate the Lyα heating of hydrogen gas with underlying perturbations in the density, temperature, and baryonic velocity. We assume that these perturbations are small and consider their contribution only to linear order. We find that, as a consequence of perturbations, the gas heating rate can be altered by a few percent compared to the heating rate in a homogeneous medium. Of particular interest are perturbations with scales comparable to or smaller than the diffusion length of Lyα photons. For perturbations on these scales, photons can interact with hydrogen atoms located in regions with different properties than where the photons originated from, changing the gas heating rate. This process can therefore be viewed as thermal conduction between regions of different temperatures, which could lead to the damping of perturbations. The spatial redistribution of photons depends in a complicated way not only on the properties of the gas, but also on the rate of frequency redistribution of photons since the scattering cross section (and hence spatial diffusion coefficient) varies by many orders of magnitude over the frequency range of interest. Therefore, as it is difficult to make a simple estimate of the diffusion length of photons, we solve the problem numerically. Our results show that the scale at which perturbations start to counteract the mean effect, and hence damp the perturbations, corresponds to a wavenumber of $k \sim 10^4$ Mpc^{-1} (comoving). That length scale is roughly two orders of magnitude smaller that the Jeans scale.

This chapter is structured as follows: At the beginning of Section 2.2 we introduce the notation and outline the formalism that is used in our analysis. We continue by describing the radiative transfer equations and the resulting radiation spectra. Heating rate calculations for the continuum and injected photons are described in Section 2.3. We present our results in Section 2.4, and finally discuss and summarize our conclusions in Section 2.5.

Throughout this chapter we assume the following values of the relevant cosmological parameters, obtained by the Planck Collaboration et al. (2014): $H_0 = 67.3$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.685$, $\Omega_m = 0.315$.

2.2 Formalism

Our formalism is based on following the time evolution of photon phase-space distribution, which is governed by the Boltzmann equation. The approach is similar to that developed for studying the cosmic microwave background (Ma & Bertschinger 1995), except that in our steady-state case, the nontrivial variable is frequency, rather than the time. In our calculation we neglect polarization since its effect on the radiation intensity is expected to be small and to include it in the calculation would require tracking twice as many variables.

We start the analysis by considering the phase-space density of photons of frequency ν, located at coordinate $\mathbf{x}$ and propagating in direction $\hat{n}$, given by the occupation number $f_\nu(\mathbf{x}, \hat{n})$. To simplify our equations, from now on we omit writing $(\mathbf{x}, \hat{n})$ explicitly, although we assume such dependence in calculations. The occupation number f_ν consists of two parts: the mean isotropic part $\overline{f}_\nu$ and direction-dependent perturbations δf_ν:

$$f_\nu = \overline{f}_\nu + \delta f_\nu \, . \tag{2.1}$$

The scale of perturbation is determined by its wavenumber k. In the equations given throughout this chapter, k is used to denote the physical wavenumber, rather than comoving, which simplifies expressions. We convert to comoving wavenumber only at the end when we report the final results and present them in figures. We assume that all perturbations are small and linear, so that we can treat them independently, since in linear perturbation theory, different k-modes are decoupled form each other. The contribution of a single k-mode to δf_ν can be expanded in a series of Legendre polynomials with coefficients $\delta f_{l\nu}$:

$$\delta f_\nu = \sum_{l=0}^{\infty} i^l (2l+1) \delta f_{l\nu} P_l(n_3) e^{ikx_3}, \tag{2.2}$$

where n_3 is the projection of unit vector $\hat{n}$ onto x_3 axis.

The IGM can be described by its mean number density $\overline{n}$ and the mean gas kinetic temperature $\overline{T}$. However, for an inhomogeneous medium, the density and temperature fields are given by:

$$n(\mathbf{x}) = \overline{n}(1 + \delta_n e^{ikx_3}) \tag{2.3}$$

and

$$T(\mathbf{x}) = \overline{T}(1 + \delta_T e^{ikx_3}) \, , \tag{2.4}$$

where δ_n and δ_T are dimensionless parameters describing the amplitudes of density and temperature perturbations, respectively.

We treat the photon field in the rest frame of the baryons–not the comoving frame–because a photon will resonantly scatter with an atom if the photon frequency matches the resonant frequency in the atom's rest frame. In that frame the overall mean velocity of atoms vanishes. We introduce linear perturbations in the baryonic velocity:

$$\mathbf{v}(\mathbf{x}) = \delta_v e^{ikx_3}\mathbf{e}_3 \ , \tag{2.5}$$

where δ_v is taken to be imaginary. Velocity divergence is then given by

$$\Theta(\mathbf{x}) = \nabla\mathbf{v}(\mathbf{x}) = ik\delta_v e^{ikx_3} = \delta_\Theta e^{ikx_3}. \tag{2.6}$$

2.2.1 Radiative Transfer

Time evolution of the photon distribution function is governed by the Boltzmann equation

$$\frac{\partial f_\nu}{\partial t} + \frac{d\nu}{dt}\frac{\partial f_\nu}{\partial \nu} + \frac{dx_i}{dt}\frac{\partial f_\nu}{\partial x_i} + \frac{dn_i}{dt}\frac{\partial f_\nu}{\partial n_i} = \left.\frac{\partial f_\nu}{\partial t}\right|_{coll} . \tag{2.7}$$

The left-hand side of the equation describes the free steaming of photons, and the collision term is on the right-hand side. To keep our analysis linear in small quantities, we ignore the last term on the left, representing gravitational lensing, because both factors are of the first order in perturbations, making the entire term second order. Therefore, our linearized collisionless equation is

$$\frac{df_\nu}{dt} \approx \frac{\partial f_\nu}{\partial t} + \frac{d\nu}{dt}\frac{\partial f_\nu}{\partial \nu} + \frac{dx_i}{dt}\frac{\partial f_\nu}{\partial x_i} \ . \tag{2.8}$$

Next, we evaluate different terms of this equation in the Newtonian gauge. The second term on the right side includes the time change in the photon frequency due to the expansion of the Universe and the relative motion of the baryons because the frequency in equations (2.7) and (2.8) is defined relative to the baryons, rather than to an observer fixed in Newtonian coordinates. We ignore contributions of the time derivative of metric perturbation because the metric potential is negligible compared to the subhorizon perturbations in the baryons that we are considering in this analysis[6]. The third term is proportional to the gradient of f_ν and the

[6]Using the basic equations of the linear perturbation theory, it can be shown that the amplitude of baryonic perturbations is proportional to $(kc/H)^2\,\Phi$, where Φ is the metric potential. For subhorizon perturbation, the wavenumber k is much larger than H/c, making Φ negligible compared to perturbations in the baryons. Similarly, it can be shown that the baryonic velocity is proportional to $(kc/H)\Phi$. Therefore, we can neglect the gravitational redshift/blueshift because it is small compared to the redshift/blueshift due to peculiar motions of baryons

only contribution to that term comes only from the perturbative part of the photon distribution function δf_ν. The accompanying factor is just the velocity of photons in the direction of the x_3 axis, which is equal to cn_3.

Focusing for now only on the collisionless Boltzmann equation, we set it equal to zero and get the following expression for the time evolution of the photon occupation number in the free-streaming (collisionless) case:

$$\left.\frac{\partial f_\nu}{\partial t}\right|_{\rm fs} = \left(H\nu + n_3^2 \nu e^{ikx_3}\delta_\Theta\right)\frac{\partial f_\nu}{\partial \nu} - ikcn_3 \sum_{l=0}^{\infty} i^l (2l+1)\delta f_{l\nu} P_l\left(n_3\right) e^{ikx_3}\,, \tag{2.9}$$

where H is the Hubble parameter.

Using the basic properties and recurrence relations of Legendre polynomials, we find the expressions for each multipole order

$$f_{l\nu} = \frac{1}{2i^l}\int f_\nu P_l(n_3) dn_3 \tag{2.10}$$

and its time derivative

$$\begin{aligned}\left.\dot{f}_{l\nu}\right|_{\rm fs} &= H\nu\left(\frac{\partial \overline{f}_{l\nu}}{\partial \nu} + e^{ikx_3}\frac{\partial \delta f_{l\nu}}{\partial \nu}\right) + \nu\delta_\Theta e^{ikx_3}\frac{l(l-1)}{(2l-1)(2l+1)}\frac{\partial \overline{f}_{(l-2)\nu}}{\partial \nu} \\ &+ \nu\delta_\Theta e^{ikx_3}\frac{(l+1)^2(2l-1)+l^2(2l+3)}{(2l-1)(2l+1)(2l+3)}\frac{\partial \overline{f}_{l\nu}}{\partial \nu} + \nu\delta_\Theta e^{ikx_3}\frac{(l+2)(l+1)}{(2l+1)(2l+3)}\frac{\partial \overline{f}_{(l+2)\nu}}{\partial \nu} \\ &- \frac{kce^{ikx_3}}{2l+1}\left[l\delta f_{(l-1)\nu} - (l+1)\delta f_{(l+1)\nu}\right]\,. \end{aligned} \tag{2.11}$$

Since we assume that $\overline{f}_\nu$ is isotropic, only the monopole term ($\overline{f}_{0\nu}$) is nonzero and the above expression is therefore greatly simplified for most multipole orders. More specifically, the second term on the right-hand side of equation 2.11 is nonzero only for $l = 2$, the third term contributes only to the equation for $l = 0$, and the fourth term vanishes for all values of l.

The right side of the full Boltzmann equation (equation 2.7) describes the change in the photon occupation number due to collisions with atoms. It consists of two terms, one describing photons scattered into the phase-space element of interest and the other describing the outgoing photons (Rybicki & dell'Antonio 1994):

$$\left.\frac{\partial f_\nu}{\partial t}\right|_{\rm coll} = \int n_H \sigma\left(\nu'\right) c f_{\nu'}(\hat{n}') R(\nu'\hat{n}', \nu\hat{n}) d\nu' d^2\hat{n}' - n_H\sigma\left(\nu\right) c f_\nu\left(\hat{n}\right), \tag{2.12}$$

where n_H is the number density of hydrogen atoms and $\sigma(\nu) = \sigma_0\Phi(\nu)$ is the collisional cross-section at frequency ν given by the cross-section at the line center σ_0 and the Voigt profile $\Phi(\nu)$:

$$\sigma(\nu) = \frac{\pi e^2}{m_e c}\frac{f_{12}}{\Delta\nu_D}\frac{a}{\pi^{3/2}}\int_{-\infty}^{+\infty} dy \frac{e^{-y^2}}{(x-y)^2 + a^2}\,, \tag{2.13}$$

where $a = A_{21}/(8\pi\Delta\nu_D)$ is the Voigt parameter and $A_{21} = 6.25 \times 10^8$ s^{-1} is the Einstein coefficient of spontaneous emission for the Lyα transition. The Doppler width of the line is given by

$$\Delta\nu_D = \nu_\alpha\sqrt{\frac{2k_BT}{m_Hc^2}} \tag{2.14}$$

and x is used to denote offset from the line center

$$x = \frac{\nu - \nu_\alpha}{\Delta\nu_D} = \frac{\Delta\nu}{\Delta\nu_D}\,. \tag{2.15}$$

The probability that a photon of frequency ν', propagating in the direction of $\hat{n}'$, will be redistributed upon scattering into a photon of frequency ν, propagating in the direction of $\hat{n}$ is represented by $R(\nu'\hat{n}', \nu\hat{n})$. It can be decomposed into a series of Legendre polynomials in terms of the scattering angle, the cosine of which is given by the dot product of $\hat{n}$ and $\hat{n}'$:

$$R(\nu'\hat{n}', \nu\hat{n}) = \frac{1}{4\pi}\sum_l R(l;\nu,\nu')P_l(\hat{n}\cdot\hat{n}')\,. \tag{2.16}$$

Plugging this into equation (2.12) and using the obtained expression in the time derivative of equation (2.10) gives

$$\begin{aligned}\left.\frac{\partial f_{l\nu}}{\partial t}\right|_{\rm coll} &= \frac{n_Hc}{2i^l}\int \sigma(\nu')\, f_{\nu'}\, d\nu' d\phi d(\cos\theta') d(\cos\theta) P_l(\cos\theta)\frac{1}{4\pi}\sum_l R(l;\nu',\nu)P_l(\hat{n}\cdot\hat{n}') \\ &- n_H\sigma(\nu)\, c\, f_{l\nu}.\end{aligned} \tag{2.17}$$

Using the spherical harmonic addition theorem

$$P_l(\hat{n}\cdot\hat{n}') = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y^*_{lm}(\theta',\phi')Y_{lm}(\theta,\phi)\,, \tag{2.18}$$

the relation between spherical harmonics and Legendre polynomials

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}P_l^m(\cos\theta)e^{im\phi} \tag{2.19}$$

and the orthogonality of Legendre polynomials, we get

$$\left.\frac{\partial f_{l\nu}}{\partial t}\right|_{\rm coll} = \int n_H \sigma(\nu')\, c\, f_{l\nu'} R(\nu', \nu)\delta_{l0} d\nu' - n_H \sigma(\nu)\, c\, f_{l\nu}. \tag{2.20}$$

Here we assume that the emission of photons is isotropic (i.e. nonzero only for $l = 0$), which is a reasonable assumption for a medium that is optically thick at the resonant frequency. The outgoing part of the collision term is direction-dependent in the case of an inhomogeneous medium. Hence, we keep that term for all multipole orders l. For multipoles with $l > 0$, this term dominates and causes their attenuation. For $l = 0$, on the other hand, the incoming and outgoing terms nearly cancel, which is why the monopole equation needs to be treated differently.

The redistribution function $R(\nu', \nu)$ is generally very complicated. However, it can be simplified if the radiation spectrum changes smoothly on the scale of a typical change of the photon frequency in a single scattering, which is on the order of $\Delta\nu_D$. If this condition is satisfied, we can use the Fokker-Planck approximation in which scattering is treated as diffusion in frequency space. Using the result of Rybicki (2006), the collision term for the monopole order then becomes

$$\frac{\partial f_{0\nu}^{coll}}{\partial t} = \frac{1}{\nu^2}\frac{\partial}{\partial \nu}\left[\nu^2 D_\nu \left(\frac{\partial f_{0\nu}}{\partial \nu} + \frac{h f_{0\nu}}{k_B T}\right)\right] + \Psi, \tag{2.21}$$

where Ψ is the photon source term describing photons injected with a frequency distribution that can be approximated by a delta function around the Lyα frequency. This term is used for describing the injected photons, whereas it vanishes in the case of the continuum photons. The parameter D_ν (in units of Hz^2 s^{-1}) is the frequency diffusivity, given by Hirata (2006):

$$D_\nu = \frac{3k_B T}{m_H}\gamma n_H x_{HI} c\Phi(\nu)\,. \tag{2.22}$$

Here $\gamma = 50$ MHz is the half width at half maximum of the Lyα resonance, x_{HI} is the neutral fraction of hydrogen, and m_H is the mass of the hydrogen atom.

Equating the result for the free streaming and the collision term, and assuming that a steady state ($\partial f_\nu/\partial t = 0$) has been reached, gives the full expression for the radiation transport. We can write an equation for each multipole order separately, producing an infinite series of coupled differential equations–the Boltzmann hierarchy. Equations for a few lowest orders are given in Appendix 2.A. In order to numerically solve this system of equations, we need to choose the highest multipole order l_{max} at which to close the hierarchy. Terminating the hierarchy at some finite order carries a risk

of transferring artificial power back to lower multipoles (Ma & Bertschinger 1995; Hu et al. 1995). We tested our results for a number of different boundary conditions for $l_{max} \sim 10$, and found that the solutions for the lowest orders ($l = 0$ and $l = 1$) are almost insensitive to the change in the boundary condition, so we choose to set $\delta f_{l_{max}+1} = 0$.

2.2.2 Unperturbed Background Solution

To obtain the mean background solution $\overline{f}_{0\nu}$ we solve the unperturbed monopole equation (i.e., the equation for $l = 0$ without any perturbative terms). For the continuum photons the unperturbed equation is

$$H\nu_\alpha \frac{\partial \overline{f}_{0\nu}}{\partial \nu} + \frac{1}{\nu_\alpha^2}\frac{\partial}{\partial \nu}\left[\nu^2 D_\nu \left(\frac{\partial \overline{f}_{0\nu}}{\partial \nu} + \frac{h}{k_B T}\overline{f}_{0\nu}\right)\right] = 0 \,. \tag{2.23}$$

In solving this equation we follow the procedure described in Chen & Miralda-Escudé (2004). The resulting spectrum (Figure 2.1, top panel), normalized to the intensity of photons on the blue side far away from the line center, shows an asymmetric absorption feature around the Lyα frequency. The shape of the feature is determined by the photon drift and diffusion in frequency caused by scattering off of hydrogen atoms. This suppression in the radiation spectrum remains fixed once a steady state has been reached; it does not redshift away with the expansion of the Universe, indicating energy transfer from the radiation field to the gas.

As shown in the top panel of Figure 2.1, the absorption feature is deeper for the gas of higher mean density (shown in blue dash-dotted line) because the scattering rate increases if there are more hydrogen atoms present. The feature is shallower for the gas of higher mean temperature (green dashed line) because in that case the energy transferred via recoils makes a smaller fraction of the average kinetic energy of the atoms. Similarly, the absorption feature is shallower for the case of non-zero atomic velocity divergence (red dotted line), which can be thought of as a bulk contribution to the kinetic energy of atoms in addition to their thermal motion. This explains why the feature changes in the same way as for an increase in temperature.

For the injected photons the background equation has an extra term $\Psi = H\nu_\alpha \delta(\nu_\alpha)$, resulting in a different spectral shape (Figure 2.1, bottom panel). If there was no scattering, the photons would be injected at the Lyα frequency, and they would simply redshift to lower frequencies, creating a spectrum shaped as a step function. However, diffusion in frequency induced by scattering transfers some of the photons from the red side of the line to the blue side. This transfer is enhanced for the gas

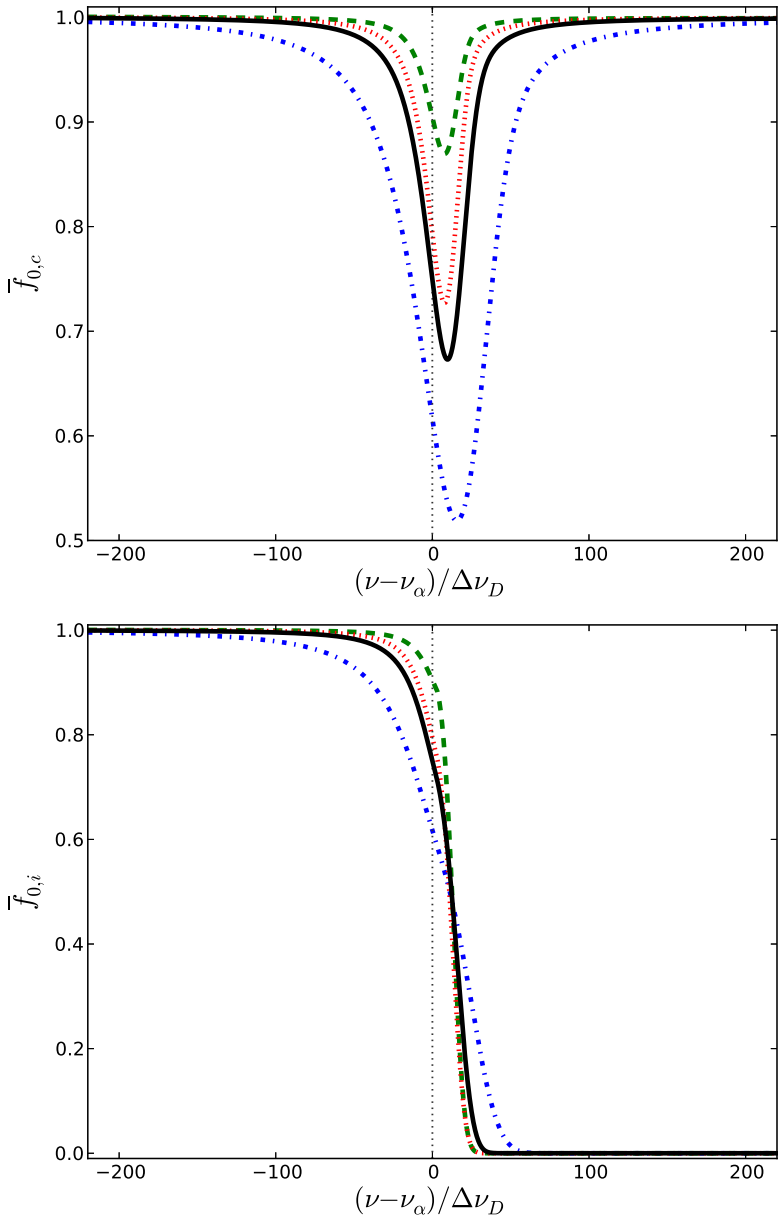

Figure 2.1: Top: Background radiation spectrum of the continuum photons around the Lyα frequency for the mean (unheated) temperature of $\overline{T}$ = 10 K and the mean density at z = 20 (solid black line), normalized to the intensity of photons far away from the Lyα frequency. The spectrum shows an asymmetric absorption feature that results from combined contributions of scattering diffusivity and atomic recoil. Additional lines show how changing the conditions in the gas can modify the absorption feature: the green dashed line represents the solution for five times higher mean temperature, the dash-dotted blue line is obtained for five times higher mean density and the red dotted line corresponds to the gas of unaltered mean temperature and density, but with a significant velocity divergence. Bottom: Same as in the top panel, but for the case of the injected photons.

of higher mean density due to an increased scattering rate. Increasing the mean temperature of the gas and introducing bulk motions have the opposite effect, as in the case of the continuum photons. Injected photons cause cooling of the gas, as the upscattering of photons to the blue side extracts energy from the gas.

2.2.3 Perturbations

Perturbative terms in our equations have one of the following elements: perturbations to the photon distribution function $\delta f_{l\nu}$, non-zero velocity divergence δ_Θ, or perturbations to the diffusivity parameter δ_D, which include density and temperature perturbations, δ_n and δ_T, respectively. The full system of coupled differential equations including perturbative terms up to linear order is given in Appendix 2.A. We numerically solve it to obtain spectra of perturbations for all multipole orders of interest. More details on the numerical implementation can be found in Appendix 2.B. Figure 2.2 shows the lowest two orders ($l = 0$ and $l = 1$) in the expansion of the photon distribution function for different types of perturbations. Spectral features that arise for perturbations with wavenumbers in the range considered in this study have characteristic widths that are on the order of several $\Delta\nu_D$ or greater, justifying the use of the Fokker-Planck approximation.

The resulting spectrum for δf_0 represents photons that are added (or subtracted, depending on whether δf_0 is positive or negative) to the mean solution $\overline{f}_0$ because of a small change in the gas temperature, density, or velocity. Since we are considering the photon phase space density in the frame of the gas, the dipole term δf_1 represents the photon flux into or out of a Lagrangian region of interest. It vanishes at the line center ($\nu \approx \nu_\alpha$) due to a very small mean free path of photons near the resonant frequency.

2.3 Heating Rates

In the case of the continuum photons, the gas and the radiation field form a closed system whose energy is conserved. The rate at which the gas is heated thus equals the negative time change of the radiation energy:

$$\left[\frac{\partial U_{gas}}{\partial t} + \frac{\partial U_{rad}}{\partial t}\right]_c = 0\,, \tag{2.24}$$

where U is used to denote the energy density (in erg cm^{-3}) of the gas and radiation. The subscript c indicates that this equation holds for the continuum photons. For the injected photons, however, there is an external source of energy that needs to be

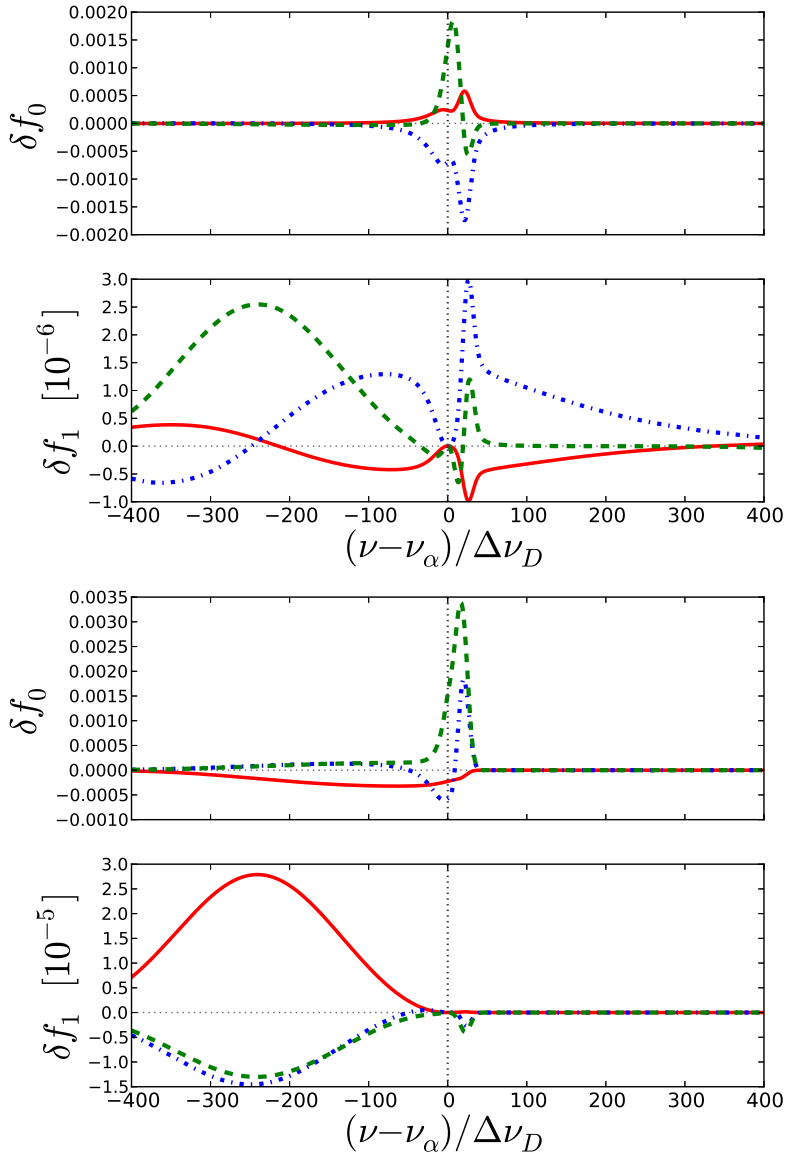

Figure 2.2: Monopole and dipole terms of the perturbed radiation field for the continuum (two upper panels) and injected (two lower panels) photons, shown with the same normalization as in Figure 2.1. Different curves show the results obtained for different types of perturbations, all with the wavenumber $k = 1$ cMpc^{-1}: green dashed curves correspond to 1% temperature perturbations ($\delta_T = 0.01$), blue dash-dotted lines represent 1% perturbations in the density ($\delta_n = 0.01$) and the red solid lines show the results for introducing 1% perturbation in the velocity divergence ($\delta_\Theta/H = 0.01$) for the continuum photons and 10 times smaller perturbation in size and amplitude for the case of the injected photons.

taken into account:

$$\left[\frac{\partial U_{gas}}{\partial t} + \frac{\partial U_{rad}}{\partial t}\right]_i = h\nu_\alpha \dot{N}_i \ , \tag{2.25}$$

where $\dot{N}_i$ is the generation rate of the injected photons.

The radiation energy density is given by

$$U_{rad} = \int n_\nu h\nu d\nu \ . \tag{2.26}$$

The number density of photons of frequency ν is n_ν (in units of cm^{-3} Hz^{-1}). It is related to the specific intensity J_ν (intensity by the number of photons, not their energy, given in units of cm^{-2} s^{-1} Hz^{-1} sr^{-1}) through the following relation:

$$n_\nu = \frac{4\pi J_\nu}{c} = \frac{8\pi\nu^2 f_{0\nu}}{c^3} \ . \tag{2.27}$$

Thus the photon energy density takes the form

$$U_{rad} = \frac{8\pi h}{c^3}\int \nu^3 f_{0\nu} d\nu \ , \tag{2.28}$$

where the integral needs to be taken over a wide enough range around the Lyα frequency to include all significant spectral features. The gas heating rate per unit volume, given in units of erg cm^{-3} s^{-1}, is the rate of change of the gas energy density $\Gamma \equiv \partial U_{gas}/\partial t$. For the sake of brevity, from now on we refer to Γ simply as the heating rate. For the continuum photons, the heating rate is given by

$$\Gamma_c = -\left.\frac{\partial U_{rad}}{\partial t}\right|_c = -\frac{8\pi h}{c^3}\int \nu^3 \left.\frac{\partial f_{0\nu}}{\partial t}\right|_c d\nu \ . \tag{2.29}$$

For the injected photons there is an additional term

$$\Gamma_i = -\frac{8\pi h}{c^3}\int \nu^3 \left.\frac{\partial f_{0\nu}}{\partial t}\right|_i d\nu + h\nu_\alpha \dot{N}_i. \tag{2.30}$$

In the first term on the right side we can make use of the expression for the time derivative of $f_{0\nu}$ given in equation 2.21. Note that in the case of the injected photons, the source function Ψ is non-zero. The photon injection rate in equation 2.30 can be written as

$$\frac{\partial N_i}{\partial t} = \frac{\partial}{\partial t}\left(\int n_{i,\nu} d\nu\right) = \frac{\partial}{\partial t}\left(\int \frac{8\pi\nu^2}{c^3} f_i d\nu\right) = \int \frac{8\pi\nu^2}{c^3}\frac{\partial f_i}{\partial t} d\nu = \int \frac{8\pi\nu^2}{c^3}\Psi d\nu \ . \tag{2.31}$$

The two terms containing the source function cancel out–the injection of photons does not contribute to the gas heating rate. The heating of the gas is caused solely by the frequency diffusivity part of the collision term given by equation (2.21).

The largest contribution to the gas heating rate comes form the part of the spectrum around the line center, thus it is convenient to separate the radiation energy density in the following way:

$$U = \frac{8\pi h}{c^3}\left[\nu_\alpha \int_{\nu_1}^{\nu_2} \nu^2 f_{0\nu} d\nu + \int_{\nu_1}^{\nu_2} (\nu - \nu_\alpha)\nu^2 f_{0\nu} d\nu\right]. \tag{2.32}$$

The corresponding heating rate is

$$\Gamma = -\frac{8\pi h}{c^3}\left[\nu_\alpha \int_{\nu_1}^{\nu_2} \nu^2 \frac{df_{0\nu}^{coll}}{dt} d\nu + \int_{\nu_1}^{\nu_2} (\nu - \nu_\alpha)\nu^2 \frac{df_{0\nu}^{coll}}{dt} d\nu\right]. \tag{2.33}$$

Using the expression given in equation 2.21, the first term on the right side becomes

$$-\frac{8\pi h\nu_\alpha}{c^3}\int_{\nu_1}^{\nu_2} \frac{\partial}{\partial\nu}\left[\nu^2 D_\nu\left(\frac{\partial f_{0\nu}}{\partial\nu} + \frac{h}{k_B T} f_{0\nu}\right)\right] d\nu = -\frac{8\pi h\nu_\alpha}{c^3}\left[\nu^2 D_\nu\left(\frac{\partial f_{0\nu}}{\partial\nu} + \frac{h}{k_B T} f_{0\nu}\right)\right]\Bigg|_{\nu_1}^{\nu_2}. \tag{2.34}$$

The contribution of this term to the total heating rate is vanishingly small because D_ν approaches zero far from the line center. The remaining term is

$$\Gamma = -\frac{8\pi h}{c^3}\int_{\nu_1}^{\nu_2} (\nu - \nu_\alpha)\frac{\partial}{\partial\nu}\left[\nu^2 D_\nu\left(\frac{\partial f_{0\nu}}{\partial\nu} + \frac{h}{k_B T} f_{0\nu}\right)\right] d\nu\,. \tag{2.35}$$

We can separate the heating rate into the contribution of the mean background radiation field and the contribution of the perturbations with

$$\Gamma = \overline{\Gamma} + e^{ikx_3}\delta\Gamma\,, \tag{2.36}$$

where $\overline{\Gamma}$ and $\delta\Gamma$ represent the background and perturbation heating, respectively. Making use of equation 2.23 for the background heating rate, and an analogous expression for the case of perturbations, obtained from equation 2.11, we get

$$\begin{aligned}
\overline{\Gamma} &= \frac{8\pi h}{c^3}\int_{\nu_1}^{\nu_2} (\nu - \nu_\alpha) H\nu_\alpha^3 \frac{\partial \overline{f}_{0\nu}}{\partial\nu} d\nu \\
&= -\frac{8\pi h\nu_\alpha^3}{c^3} H(\nu - \nu_\alpha)(\overline{f}_\alpha - \overline{f}_{0\nu})\Big|_{\nu_1}^{\nu_2} + \frac{8\pi h\nu_\alpha^3}{c^3} H\int_{\nu_1}^{\nu_2} (\overline{f}_\alpha - \overline{f}_{0\nu}) d\nu\,, \qquad (2.37)\\
\delta\Gamma &= \frac{8\pi h}{c^3}\int_{\nu_1}^{\nu_2}\left(H\nu_\alpha^3\frac{\partial\delta f_{0\nu}}{\partial\nu} + \frac{\nu_\alpha^3\delta_\Theta}{3}\frac{\partial\overline{f}_{0\nu}}{\partial\nu} + \nu_\alpha^2 kc\delta f_{1\nu}\right)(\nu - \nu_\alpha) d\nu \\
&= \frac{8\pi h\nu_\alpha^3}{c^3} H\left[(\nu - \nu_\alpha)\delta f_\alpha\Big|_{\nu_1}^{\nu_2} - \int_{\nu_1}^{\nu_2}\delta f_{0\nu} d\nu\right] - \frac{8\pi h\nu_\alpha^3\delta_\Theta}{3c^3}(\nu - \nu_\alpha)\left(\overline{f}_\alpha - \overline{f}_{0\nu}\right)\Big|_{\nu_1}^{\nu_2} \\
&\quad + \frac{8\pi h\nu_\alpha^3\delta_\Theta}{3c^3}\int_{\nu_1}^{\nu_2}\left(\overline{f}_\alpha - \overline{f}_{0\nu}\right) d\nu + \frac{8\pi h\nu_\alpha^2}{c^3} kc\int_{\nu_1}^{\nu_2}(\nu - \nu_\alpha)\delta f_{1\nu} d\nu\,. \qquad (2.38)
\end{aligned}$$

The above formulae are appropriate for frequencies around the line center. However, they might cause significant numerical errors in the wings of the line due to approximations made in deriving them. Therefore, we use another expression to evaluate heating rate in the wings:

$$\Gamma = -\frac{8\pi h}{c^3}\left[(\nu-\nu_\alpha)\nu^2 D_\nu\left(\frac{\partial f_{0\nu}}{\partial\nu}+\frac{h}{k_B T}f_{0\nu}\right)\Big|_{\nu_1}^{\nu_2}\right]+\frac{8\pi h}{c^3}\left[\int_{\nu_1}^{\nu_2}\nu^2 D_\nu\left(\frac{\partial f_{0\nu}}{\partial\nu}+\frac{h}{k_B T}f_{0\nu}\right)d\nu\right] . \tag{2.39}$$

We report the calculated heating rates in terms of a dimensionless quantity, which we call the *relative heating*, that measures the energy transferred to the gas per Hubble time, relative to the thermal energy of the gas ($3\overline{n}k_B\overline{T}/2$):

$$\frac{\Gamma}{\frac{3}{2}\overline{n}k_B\overline{T}H(z)} = \frac{\overline{\Gamma}}{\frac{3}{2}\overline{n}k_B\overline{T}H(z)} + \frac{\delta\Gamma e^{ikx_3}}{\frac{3}{2}\overline{n}k_B\overline{T}H(z)} \quad , \tag{2.40}$$

where $\overline{n}$ is the number density of all baryons, not just hydrogen atoms. Contributions of density, temperature and velocity perturbations to the heating rate are incorporated into $\delta\Gamma$ and can be treated independently for each type of perturbations:

$$\delta\Gamma \;=\; \frac{\partial\Gamma}{\partial n}\delta_n + \frac{\partial\Gamma}{\partial T}\delta_T + \frac{\partial\Gamma}{\partial\Theta}\delta_\Theta \quad . \tag{2.41}$$

The perturbative part of equation 2.40 is given by

$$\frac{\delta\Gamma e^{ikx_3}}{\frac{3}{2}\overline{n}k_B\overline{T}H(z)} = \left[C_n\delta_n + C_T\delta_T + C_\Theta\frac{\delta_\Theta}{H}\right]\frac{\overline{J}_\alpha}{\tilde{J}_0}e^{ikx_3} \quad , \tag{2.42}$$

where $\overline{J}_\alpha$ is the specific intensity of incoming photons and

$$\tilde{J}_0 = \frac{n_H c}{4\pi\nu_\alpha} = \frac{2\nu_\alpha^2\tilde{f}_0}{c^2} \tag{2.43}$$

is the intensity corresponding to one photon per frequency octave per hydrogen atom in the Universe (Chen & Miralda-Escudé 2004). In terms of energy intensity, this corresponds to $\tilde{J}_0 h\nu_\alpha \approx 2.5\times10^{-20}$ erg cm^{-2} s^{-1} sr^{-1} Hz^{-1} at $z = 20$. In the model of Ciardi & Madau (2003), the intensity of Lyα background at $z \sim 20$ is on the order of 10^{-20} erg cm^{-2} s^{-1} sr^{-1} Hz^{-1}, making $\overline{J}_\alpha/\tilde{J}_0$ a factor of order unity. In general, one expects it to be a rapidly increasing function of redshift. Since it takes ≥ 1 H-ionizing ($\nu > \frac{4}{3}\nu_\alpha$) photons per atom to ionize the Universe, and since the non-ionizing photons that redshift into Lyman series lines do not suffer from absorption in the emitting galaxies, we expect that $\overline{J}_\alpha/\tilde{J}_0$ should reach unity at an early stage of reionization (Chen & Miralda-Escudé 2004).

We have defined dimensionless heating coefficients C for all three types of perturbations by

$$C_n = \frac{2\tilde{f_0}}{3\overline{n}k_B\overline{T}H}\frac{\partial\Gamma}{\partial n}, \tag{2.44}$$

$$C_T = \frac{2\tilde{f_0}}{3\overline{n}k_B\overline{T}H}\frac{\partial\Gamma}{\partial T}, \text{ and} \tag{2.45}$$

$$C_\Theta = \frac{2\tilde{f_0}}{3\overline{n}k_B\overline{T}}\frac{\partial\Gamma}{\partial\Theta}; \tag{2.46}$$

these represent the heat input per Hubble time in units of the thermal energy of the gas, if $\overline{J}_\alpha/\tilde{J}_0 = 1$.

2.4 Results and Discussion

We perform calculations described in the previous sections for neutral hydrogen gas ($x_{HI} = 1$) at the redshift of $z = 20$. The mean baryon number density is easily obtained from the current baryon density of the Universe $\overline{n}_0 = 2.5 \times 10^{-7}$ cm^{-3} (WMAP-9 result, Bennett et al. 2013) as $\overline{n}(z) = \overline{n}_0(1+z)^3$. Finally, to get the number density of hydrogen atoms we take into account that over 90% (by number) of the baryonic content is in the form of hydrogen atoms. For the mean temperature of the gas we take the value $\overline{T} = 10$ K (see Figure 1 in Pritchard & Furlanetto 2006).

2.4.1 Heating from Unperturbed Radiation

The calculated contribution to the relative heating coming from the mean (unperturbed) background photons can be expressed as

$$\frac{\overline{\Gamma}_c}{\frac{3}{2}\overline{n}k_B\overline{T}H} = 0.13\frac{\overline{J}_{\alpha,c}}{J_0} \tag{2.47}$$

for the continuum photons, and as

$$\frac{\overline{\Gamma}_i}{\frac{3}{2}\overline{n}k_B\overline{T}H} = -0.07\frac{\overline{J}_{\alpha,i}}{J_0} \tag{2.48}$$

for the injected photons. The relative heating caused by the scattering of the injected photons is negative, indicating that the gas is cooled down by this interaction. At this temperature and density, heating of the gas caused by the scattering of the continuum photons prevails over cooling by the injected photons, not only because the effect itself is slightly stronger, but also because the flux of the injected photons is smaller than the flux of the continuum photons; the ratio of the injected and continuum photons is around 10-20% (Pritchard & Furlanetto 2006; Chuzhoy & Shapiro 2007).

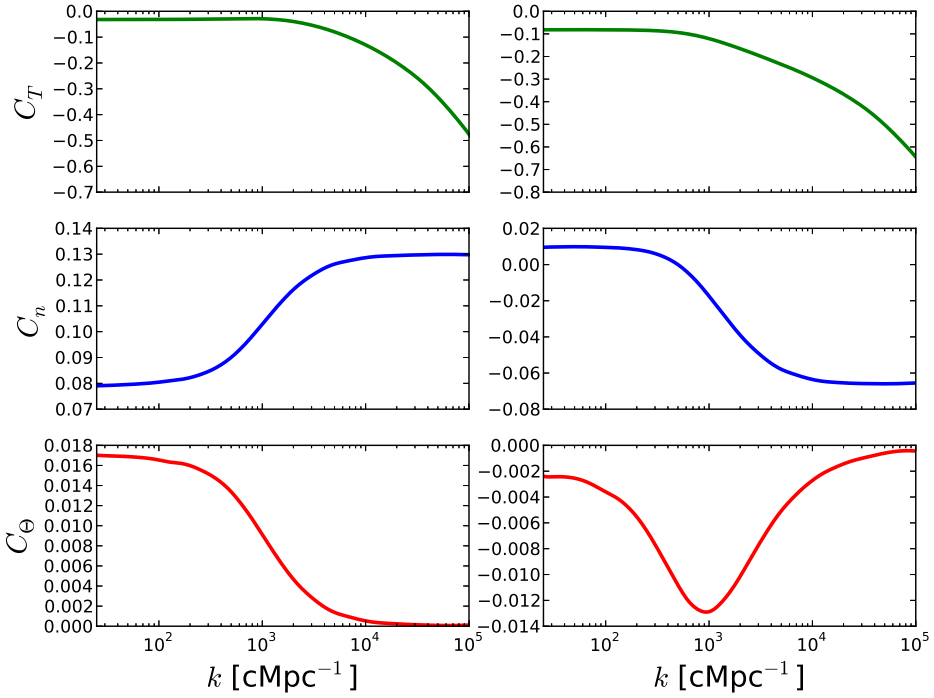

Figure 2.3: Heating coefficients, defined by expressions (2.44), (2.45), (2.46), for the continuum (left) and injected photons (right) show contributions of temperature, density and velocity perturbations (from top to bottom) to the total heating rate, as functions of the perturbation wavenumber k, given in comoving units.

2.4.2 Heating from Perturbations

A new result of this study is the additional contribution to the relative heating caused by inhomogeneities in the gas. The differential relative heating due to perturbations is given by equation 2.42. The values of all three heating coefficients C for perturbations of different wavenumbers k are shown in Figure 2.3.

On large scales, corresponding to small values of the wavenumber k, having a positive perturbation in the temperature or density is similar to having a region with no perturbations, but with an increased mean value. As shown in Figure 2.1 and discussed in Section 2.2.1, an increase in $\overline{T}$ causes the absorption feature in the spectrum of $\overline{f}_0$ to be shallower, whereas an increase in $\overline{n}$ makes the feature deeper. The heating is proportional to the integral of the difference between the spectrum without any scattering (which is just a flat spectrum for the case of the continuum photons) and the real spectrum, and hence it is proportional to the area of the absorption feature. Therefore, the heating will be smaller in the case of higher

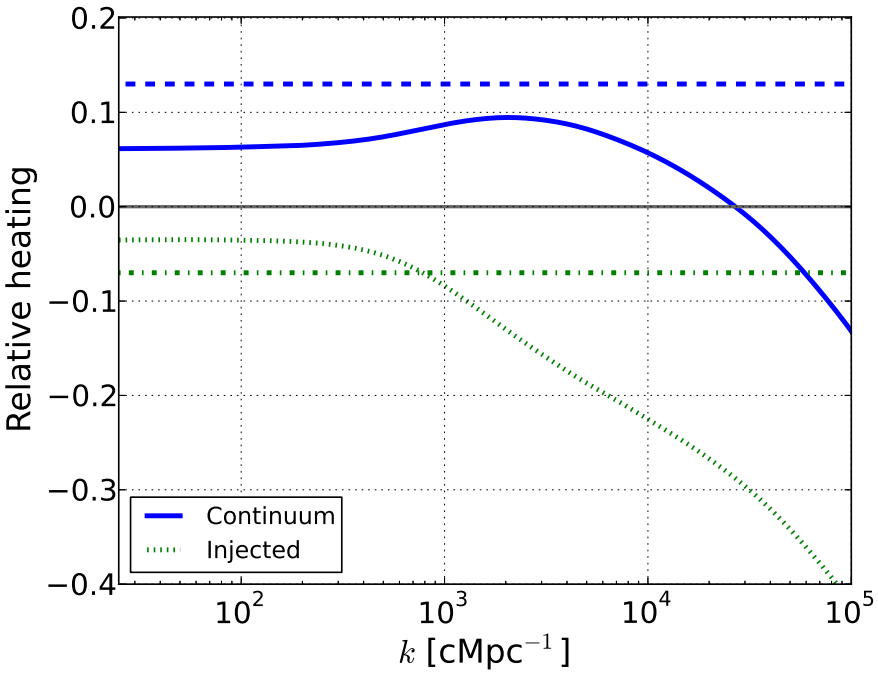

Figure 2.4: Solid and dotted lines show the joint contribution of the temperature and density perturbations to the relative heating, for the continuum and injected photons respectively, as a function of perturbation wavenumber (comoving). These values need to be multiplied by the amplitude of the density perturbations to give the perturbative heating that can then be compared to the unperturbed effect, shown in dashed and dash-dotted lines. On very small scales, the perturbative heating from the continuum photons changes sign relative to the mean effect, causing damping of small scale perturbations. The same does not occur for the injected photons – they cause negative heating (i.e. cooling) of gas on all scales. Their contribution relative to the continuum photons is diminished by the fact that their intensity is lower, making only $\sim 10-20\%$ of the intensity of the continuum photons.

$\overline{T}$, which is why C_T has negative values on large scales. On the other hand, C_n is positive, since an increase in $\overline{n}$ causes a larger heating. The absolute values of heating coefficients increase for larger values of k, indicating that these effects are enhanced on smaller scales in the case of the continuum photons.

To estimate the contribution of perturbations to the total heating of the gas, we need to know the relative strength of different kinds of perturbations, in addition to the values of the heating coefficients. Naoz & Barkana (2005) calculated the ratio of temperature and density perturbations as a function of wavenumber k at several redshifts, including $z = 20$, the epoch that we consider in this work. The ratio of δ_T and δ_n at $z = 20$ is almost constant for small-scale perturbations, $\delta_T/\delta_n \approx 0.55$.

We make use of this result to show (Figure 2.4) the joint contribution of the density and temperature perturbations ($C_n + 0.55C_T$), relative to the heating caused by the unperturbed background photons. We ignore perturbations in the velocity because the magnitude of C_Θ is much smaller than that of C_T or C_n for most wavenumbers.

On large scales (i.e., small values of k), the heating caused by perturbations has the same sign as the mean effect: positive for the continuum photons and negative for the injected photons, which indicates a small increase in the gas heating by the continuum photons in regions of higher density and temperature, due to the increased scattering rate. Cooling by the injected photons in those regions will also increase. These additional contributions to the heating are very small. The values shown in Figure 2.4, which are already a factor of ~ 2 smaller than the mean effect, need to be multiplied by the amplitude of density perturbations to give the actual relative heating. Since our formalism is based on the linear perturbation theory, it is applicable to perturbations on the order of a few percent or smaller. Therefore, the additional heating at large scales can only make a few percent of the mean (unperturbed) Lyα heating.

An interesting feature appears for perturbations on very small scales ($k \sim 2 \times 10^4$ cMpc^{-1}). For the case of the continuum photons, values of the relative heating, shown in Figure 2.4, turn from positive to negative. On length scales below that threshold, perturbations act in the opposite direction from the unperturbed effect; they reduce the heating of the gas in positively perturbed regions (i.e., regions of higher density and temperature than the mean). The opposite happens for cooler and underdense regions. The effect of the injected photons remains unchanged. Perturbations on scales smaller than that corresponding to $k \sim 2 \times 10^4$ cMpc^{-1} will be damped due to the effect of thermal conduction by Lyα photons. This length scale, however, is roughly two orders of magnitude smaller than the Jeans scale.

For gas of higher mean temperature, e.g., with $\overline{T} = 20$ K, the result stays qualitatively the same, only the values of the relative heating, both the mean effect and the contribution of perturbations, are reduced.

2.5 Conclusions

The resonant scattering of Lyα photons produced by early generations of luminous objects can cause moderate heating of high-redshift IGM. Lyα photons and hydrogen atoms exchange energy during scattering due to atomic recoil. Details of radiative transfer are further complicated by frequency diffusion of photons caused

by scattering and the drift to lower frequencies due to Hubble expansion. In the optically thick limit, the gas and the radiation field approach statistical equilibrium, which greatly reduces the energy exchange. Taken into account all these effects, an asymmetric absorption feature is created in the radiation spectrum around the Lyα frequency in the case of the continuum photons which redshift directly into Lyα from the blue side of the line. Photons that redshift into higher resonances and then cascade into Lyα are called the injected photons. Their spectrum has a shape of a modified step function around the Lyα frequency. Scattering of the continuum photons causes heating of the IGM proportional to the area of the absorption feature. This heating is higher for gas of lower mean temperature and higher density. The injected photons, on the other hand, cause cooling of the gas because they preferentially scatter off atoms moving in the opposite direction, and hence slow them down.

We study the effect of Lyα scattering on high-redshift ($z = 20$) IGM with linear perturbations in density, temperature, and velocity divergence. We are primarily interested in small-scale perturbations that can be affected by thermal conduction via Lyα photons. For perturbations with scales smaller than the Lyα diffusion length-scale, photons can diffuse into regions where they are further away from being in a statistical equilibrium with the gas, causing enhancement in the energy exchange. To find the exact scale at which this occurs, we solve radiative transfer equations numerically, using the Fokker-Planck approximation.

We find that the scale at which this effect becomes relevant is very small, corresponding to a comoving wavenumber of $k \sim 2 \times 10^4$ Mpc^{-1}, which is a factor of ~ 100 smaller than the Jeans scale. On larger scales, where structures in the IGM are expected to be present, the heating perturbations add a correction to the mean effect that is on the order of the amplitude of the density or temperature perturbation in the gas. Since our formalism is based on the linear perturbation theory, this makes only a few percent difference to the gas heating in typical cases.

APPENDIX

2.A System of Equations

Equations for different multipoles of the perturbed radiation field are given by:

$$\begin{aligned} H\nu\frac{\partial\delta f_{0\nu}}{\partial\nu} &+ \frac{1}{3}\nu\delta_\Theta\frac{\partial\overline{f}_{0\nu}}{\partial\nu} + kc\delta f_{1\nu} + \frac{\partial\overline{D}_\nu}{\partial\nu}\left(\frac{\partial\delta f_{0\nu}}{\partial\nu} + \frac{h}{k_BT}\delta f_{0\nu}\right) \\ &+ \frac{\partial\overline{D}_\nu}{\partial\nu}\delta_D\left(\frac{\partial\overline{f}_{0\nu}}{\partial\nu} + \frac{h}{k_BT}\overline{f}_{0\nu}\right) + \overline{D}_\nu\left(\frac{\partial^2\delta f_{0\nu}}{\partial\nu^2} + \frac{h}{k_BT}\frac{\partial\delta f_{0\nu}}{\partial\nu}\right) \\ &+ \overline{D}_\nu\delta_D\left(\frac{\partial^2\overline{f}_{0\nu}}{\partial\nu^2} + \frac{h}{k_BT}\frac{\partial\overline{f}_{0\nu}}{\partial\nu}\right) = 0\ , \quad (l=0) \end{aligned} \tag{2.49}$$

$$H\nu\frac{\partial\delta f_{1\nu}}{\partial\nu} - \frac{kc}{3}\left(\delta f_{0\nu} - 2\delta f_{2\nu}\right) - n_H c\sigma(\nu,T)\delta f_{1\nu} = 0\ , \quad (l=1) \tag{2.50}$$

$$H\nu\frac{\partial\delta f_{2\nu}}{\partial\nu} - \frac{2}{15}\nu\delta_\Theta\frac{\partial\overline{f_{0\nu}}}{\partial\nu} - \frac{kc}{5}\left(2\delta f_{1\nu} - 3\delta f_{3\nu}\right) - n_H c\sigma(\nu,T)\delta f_{2\nu} = 0\ , \quad (l=2) \tag{2.51}$$

$$\cdots$$

$$\begin{aligned} H\nu\frac{\partial\delta f_{l_{max}\nu}}{\partial\nu} &- \frac{kc}{2l_{max}+1}\left(l_{max}\delta f_{(l_{max}-1)\nu} - (l_{max}+1)\delta f_{(l_{max}+1)\nu}\right) \\ &- n_H c\sigma(\nu,T)\delta f_{l_{max}\nu} = 0\ , \quad (l=l_{max}), \end{aligned} \tag{2.52}$$

where k is the physical wavenumber, not comoving. In the first equation, $\overline{D}_\nu$ is used to denote the mean value of the diffusivity parameter. The full expression for this quantity, including perturbations, can be written as $D_\nu = \overline{D}_\nu(1+\delta_D) = \overline{D}_\nu(1+\delta_n+\delta_T)$.

2.B Numerical Calculations

To solve the above described system of equations, we truncate the series at $l_{max} = 8$ by setting $\delta f_{9\nu} = 0$. The choice of the largest considered multipole could of course be different. In our analysis we only use the solutions for the monopole and dipole terms, but we want to keep as many higher orders as possible to avoid introducing significant errors into the lowest mulitpoles by making an artificial truncation of the series too close to them. On the other hand, tracking too many multipole orders becomes computationally challenging. We find that keeping nine multipoles is optimal: the computation can be done in a reasonable amount of time and increasing l_{max} by one changes the resulting heating rate by 2% at most (in most cases much less than that).

Once we have a finite set of differential equations, we create an equidistant frequency grid containing a large number (7×10^5) of frequencies centered at ν_α. The frequency

range covered in our grid spans 3.3×10^{12} Hz (corresponding to $\sim 1000\ \Delta\nu_D$) on each side of the Lyα frequency. These numbers could have been chosen differently without significantly affecting the final result. For example, decreasing the frequency range by 10% changes the computed heating rate by no more than $1 - 2\%$. We can change the size of the frequency grid to test the convergence of our method. Increasing the number of frequencies in the grid spanning the same frequency range to 8×10^5 changes the result by $\sim 2\%$ or less.

In order to avoid numerical errors that can occur near the boundaries of the grid, in our calculations we do not take into account outer 3×10^4 frequencies at both ends of the grid. As mentioned in section 2.3, to calculate the heating rate we use equations (2.37) and (2.38) in the central part of the grid and equation (2.39) in the outer parts. The exact number of frequencies included in these outer parts, the so-called wings, does not significantly affect the result, as long as the central spectral features, contained within $\sim 100\Delta\nu_D$ of the Lyα line center, are not included. The difference between having $\sim 20\%$ of frequencies in the wings and having 40% is less than 4%.

For the described grid of 7×10^5 frequencies and with 9 multipoles, our system of equations forms a matrix of dimension $(9 \times 7 \times 10^5) \times (9 \times 7 \times 10^5)$. Manipulating such a large matrix can be challenging. Fortunately, most of the elements of this matrix are zero, hence we make use of SciPy sparse matrix package (scipy.sparse) to construct the matrix and solve the sparse linear system.

To approximate differentiation in the equations, were first used the central difference method

$$f'(\nu) = \frac{f(\nu + \Delta\nu) - f(\nu - \Delta\nu)}{2\Delta\nu} + O(\Delta\nu^2), \tag{2.53}$$

where $\Delta\nu$ is the grid step. However, due to numerical oscillations that occurred near the end of the frequency grid, we introduced dissipation in the form of the forward difference contribution

$$f'(\nu) = \frac{f(\nu + \Delta\nu) - f(\nu)}{\Delta\nu} + O(\Delta\nu). \tag{2.54}$$

The derivatives are therefore given by a linear combination of the central and forwards difference terms

$$f'(\nu) = \epsilon \frac{f(\nu + \Delta\nu) - f(\nu)}{\Delta\nu} + (1 - \epsilon)\, \frac{f(\nu + \Delta\nu) - f(\nu - \Delta\nu)}{2\Delta\nu}\,. \tag{2.55}$$

Parameter ϵ describes the contribution of the forward difference method, and it changes linearly from zero at the center to unity at the ends of the frequency grid.

References

Bennett, C. L., Larson, D., Weiland, J. L., et al., 2013, ApJS, 208, 20, 20

Bowman, J. D., & Rogers, A. E. E., 2010, Nature, 468, 796

Chen, X., & Miralda-Escudé, J., 2004, ApJ, 602, 1

Chuzhoy, L., & Shapiro, P. R., 2007, ApJ, 655, 843

Ciardi, B., & Madau, P., 2003, ApJ, 596, 1

Ciardi, B., & Salvaterra, R., 2007, MNRAS, 381, 1137

Dillon, J. S., Liu, A., Williams, C. L., et al., 2014, PhRvD, 89.2, 023002, 023002

Field, G. B., 1958, Proceedings of the IRE, 46, 240

Furlanetto, S. R., Oh, S. P., & Briggs, F. H., 2006, PhR, 433, 181

Furlanetto, S. R., & Pritchard, J. R., 2006, MNRAS, 372, 1093

Higgins, J., & Meiksin, A., 2009, MNRAS, 393, 949

Hirata, C. M., 2006, MNRAS, 367, 259

Hu, W., Scott, D., Sugiyama, N., & White, M., 1995, PhRvD, 52, 5498

Lonsdale, C. J., Cappallo, R. J., Morales, M. F., et al., 2009, IEEE Proceedings, 97, 1497

Ma, C.-P., & Bertschinger, E., 1995, ApJ, 455, 7

Madau, P., Meiksin, A., & Rees, M. J., 1997, ApJ, 475, 429

Meiksin, A., 2006, MNRAS, 370, 2025

Morales, M. F., & Wyithe, J. S. B., 2010, ARA&A, 48, 127

Naoz, S., & Barkana, R., 2005, MNRAS, 362, 1047

Paciga, G., Albert, J. G., Bandura, K., et al., 2013, MNRAS, 433, 639

Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J., 2012, ApJ, 753, 81, 81

Parsons, A. R., Liu, A., Aguirre, J. E., et al., 2014, ApJ, 788, 106, 106

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al., 2014, A&A, 571, A16, A16

Pritchard, J. R., & Furlanetto, S. R., 2006, MNRAS, 367, 1057

Pritchard, J. R., & Loeb, A., 2012, Reports on Progress in Physics, 75.8, 086901, 086901

Rawlings, S., & Schilizzi, R., 2011, ArXiv e-prints

Rybicki, G. B., 2006, ApJ, 647, 709

Rybicki, G. B., & dell'Antonio, I. P., 1994, ApJ, 427, 603

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al., 2013, A&A, 556, A2, A2

Wouthuysen, S. A., 1952, AJ, 57, 31

Chapter 3

GIANT CLUMPS IN THE FIRE SIMULATIONS: A CASE STUDY OF A MASSIVE HIGH-REDSHIFT GALAXY

ABSTRACT

The morphology of massive star-forming galaxies at high redshift is often dominated by giant clumps of mass $\sim 10^8 - 10^9$ $M_\odot$ and size $\sim 100 - 1000$ pc. Previous studies have proposed that giant clumps might have an important role in the evolution of their host galaxy, particularly in building the central bulge. However, this depends on whether clumps live long enough to migrate from their original location in the disk or whether they get disrupted by their own stellar feedback before reaching the center of the galaxy. We use cosmological hydrodynamical simulations from the FIRE (Feedback in Realistic Environments) project which implement explicit treatments of stellar feedback and interstellar medium physics to study the properties of these clumps. We follow the evolution of giant clumps in a massive ($M_* \sim 10^{10.8}$ $M_\odot$ at $z = 1$), disky, gas-rich galaxy from redshift $z \gtrsim 2$ to $z = 1$. Even though the clumpy phase of this galaxy lasts over a gigayear, individual gas clumps are short-lived, with mean lifetime of massive clumps of ~ 20 Myr. During that time, they turn between 0.1% and 20% of their gas into stars before being disrupted, similar to local giant molecular clouds. Clumps with $M \gtrsim 10^7$ $M_\odot$ account for ~ 20 % of the total star formation in the galaxy during the clumpy phase, producing $\sim 10^{10}$ $M_\odot$ of stars. We do not find evidence for net inward migration of clumps within the galaxy. The number of giant clumps and their mass decrease at lower redshifts, following the decrease in the overall gas fraction and star-formation rate.

This chapter has been published as Oklopčić, A., Hopkins, P. F., Feldmann, R., Kereš, D., Faucher-Giguère, C.-A., Murray, N. 2017, MNRAS, 465, 952.

3.1 Introduction

Most massive ($M_* \sim 10^{10-11}$ $M_\odot$) star-forming galaxies at high redshift ($z \sim 1-3$) have much more irregular morphology than star-forming galaxies of similar mass in the local Universe. Although these galaxies often show signs of having extended rotating disks, their structure is dominated by a few giant star-forming clumps (e.g. Cowie et al. 1995; van den Bergh et al. 1996; Genzel et al. 2011; Guo et al. 2012), especially in the UV light which traces young stars. The size of these clumps is usually between a few hundred parsecs and a kiloparsec, with mass in the range $\sim 10^8 - 10^9$ $M_\odot$. Clumps typically comprise a few percent of a galaxy's mass, but they can account for $\sim 10-20\%$ of its total star formation (Elmegreen et al. 2009; Wuyts et al. 2012; Guo et al. 2015).

Clumpy morphology of galaxies at high redshift has been established in numerous observational campaigns in terms of different tracers of star formation activity. Giant clumps have been observed in the rest-frame UV and optical images (Elmegreen et al. 2007; Elmegreen et al. 2009; Förster Schreiber et al. 2011; Guo et al. 2012), in the rest-frame optical line emission spectra (Genzel et al. 2008; Genzel et al. 2011), resolved maps of molecular gas (Tacconi et al. 2013) and in the line emission of lensed galaxies (Jones et al. 2010; Swinbank et al. 2010; Livermore et al. 2012; Livermore et al. 2015). On the other hand, maps of stellar mass distribution of those galaxies do not show very prominent clumps (Wuyts et al. 2012).

It is possible that the irregular morphology of some high-redshift galaxies is due to an ongoing merger (Somerville et al. 2001). However, the overall abundance of clumpy galaxies is too high compared to the expected merger rate to explain all of them as merging systems (Dekel et al. 2009a; Stewart et al. 2009; Hopkins & Hernquist 2010; Hopkins et al. 2010). Alternatively, clumpy morphology can be the result of disk fragmentation due to gravitational instabilities in gas-rich disks (e.g. Noguchi 1999; Dekel et al. 2009b). The observed structure and kinematics of high-redshift clumpy galaxies suggest that many of them do in fact have underlying rotating disks (e.g. Elmegreen et al. 2007; Genzel et al. 2008; Genzel et al. 2011) and that they are very gas-rich, with gas fraction $\sim 50\%$ (Tacconi et al. 2010).

Local fragmentation occurs in regions of the disk where self-gravity of gas and stars overcomes the pressure support in the form of velocity dispersion and the shearing effect of differential rotation. Disk stability can be parametrized in terms of the Toomre Q parameter (Toomre 1964) given by

$$Q = \frac{\sigma \kappa}{\pi G \Sigma} , \tag{3.1}$$

where σ is the 1D velocity dispersion, κ is the epicyclic frequency (which is related to the circular frequency $\kappa \propto \Omega$), and Σ is the mass surface density. If Q is above the critical value of order unity, the disk is stable against fragmentation. On the other hand, if $Q < 1$, the disk is locally unstable and will undergo gravitational collapse. This scenario is used to explain the formation of giant molecular clouds (GMCs) in the local Universe. Assuming that giant clumps at high redshift form via the same mechanism as GMCs, but in more gas-rich environments, can explain many of their properties, as we discuss below (see also Murray et al. 2010).

Disks with large surface densities which drive Q below the critical value can experience gravitational instabilities which lead to disk fragmentation. Fragmentation can then stir up the disk, causing velocity dispersions to increase, and hence increasing Q. If Q grows beyond the critical value, the disk becomes stable, further fragmentation is suppressed and the velocity dispersion gradually decreases, consequently bringing down the value of Q along with it. In this way disks can self-regulate and maintain a marginally stable state with $Q \approx 1$. Analysis of gravitational instability in a disk with two components (gas and stars) is slightly more complicated (see e.g. Jog 1996; Elmegreen 2011), but qualitatively similar to this single-component analysis.

The characteristic size of fragments is given by $\lambda_c \sim G\Sigma/\Omega^2$ (Binney & Tremaine 2008). For a disk with $Q \approx 1$ this becomes $\lambda_c \sim \sigma/\Omega$. The characteristic radius and mass of fragments (clumps) in that case are[1]

$$R_{cl} \propto \frac{\sigma}{\Omega} \propto \frac{G\Sigma}{\Omega^2} \propto \frac{\Sigma R^2}{M} R \propto f_g' R \tag{3.2}$$

$$M_{cl} \propto \Sigma R_{cl}^2 \propto \Sigma f_g'^2 R^2 \propto f_g'^3 M. \tag{3.3}$$

Here f_g' denotes the ratio of the gas mass to the total (baryonic and dark matter) mass of the galaxy. It is proportional to what is in the rest of the chapter referred to as the gas fraction of the galaxy – the ratio of the gas mass to the baryonic mass of the galaxy – and which we denote by f_g. R and M are the radius and the mass of the galaxy, respectively. Observations indicate that star-forming galaxies at high redshift have much higher gas fractions ($f_g \sim 0.3 - 0.7$) compared to galaxies in the local Universe ($f_g \sim 0.1$). Therefore, these relations can explain the difference in

[1] If the gas component is characterized by $Q = 1$, then $\sigma \propto G\Sigma/\Omega$, where σ and Σ are the gas velocity dispersion and the gas surface density, respectively. We use the expression for the angular frequency $\Omega = v_c/R = \sqrt{GM/R^3}$, where v_c is the circular velocity at radius R and M is the total mass enclosed within that radius. The mass of gas in the galaxy is proportional to ΣR^2; hence, the fraction of the total mass of the galaxy that is in the gas component is $f_g' \propto \Sigma R^2/M$.

size and mass of GMCs (with $R \sim 100$ pc, $M \sim 10^{5-6}$ $M_\odot$) in the Milky Way, and giant clumps in high-redshift galaxies (with $R \sim 1$ kpc, $M \sim 10^{8-9}$ $M_\odot$), assuming that they form via the same type of gravitational instability. In other words, the giant clumps observed in high-redshift galaxies seem to be a consequence of their gas-rich nature, caused by fresh supplies of gas which are continuously being provided through accretion of gas from the intergalactic medium (Kereš et al. 2005; Kereš & Hernquist 2009; Dekel et al. 2009b; Brooks et al. 2009; Faucher-Giguère et al. 2011).

The importance of giant clumps for the evolution of high-redshift galaxies is a subject of ongoing research. In recent years, it has been proposed that giant clumps may be responsible for the formation of bulges in their host galaxies, as an alternative to the classical bulge formation scenario involving mergers. In these models, clumps form throughout the disk, lose angular momentum due to dynamical friction and gravitational torques within the disk and gradually sink towards the center of the galaxy (Noguchi 1999; Immeli et al. 2004; Dekel et al. 2009b). Typical time-scales over which clumps can migrate to the center are a few times 10^8 yr (Ceverino et al. 2010; Bournaud et al. 2011), which is comparable to a few disk orbital times.

However, it is uncertain whether individual clumps can survive that long. Because they are sites of intense star formation, clumps are exposed to strong stellar feedback which could destroy them on much shorter time-scales ($10^6 - 10^8$ yr). Analytical studies of this problem give different results depending on the assumptions made (Murray et al. 2010; Krumholz & Dekel 2010; Dekel & Krumholz 2013). Over the past several years, a number of groups have used numerical simulations, both cosmological and idealized simulations of isolated galaxies, to study the phenomenon of giant clumps (e.g. Ceverino et al. 2010; Hopkins et al. 2012; Mandelker et al. 2014; Bournaud et al. 2014; Tamburello et al. 2015). In most cases, gas-rich disks with properties similar to those of observed high-redshift galaxies do break up into large clumps of mass $\sim 10^7 - 10^9$ $M_\odot$. However, what happens to the clumps afterwards–how long they live and whether they migrate to the center of the galaxy before being disrupted–heavily depends on the details of the ISM physics and stellar feedback implemented in the simulation. For example, most simulations which model stellar feedback as heating from supernovae only, while ignoring some other modes of feedback such as radiation pressure, produce massive, gravitationally bound clumps which manage to survive for hundreds of millions of years, long enough to sink to the center of the galaxy.

On the other hand, the cosmological simulations of Genel et al. (2012), which include a phenomenological model of feedback in the form of strong momentum-driven winds, have giant clumps which get disrupted on time-scales of $\sim$ 50 Myr. Improved physical models of feedback have been implemented by Hopkins et al. (2012) in simulations of isolated galaxies with properties which match those of high-redshift star-forming disk galaxies. They found that giant clumps get disrupted by stellar feedback in a few times 10^7 yr, after converting only a few percent of their gas into stars, which suppresses the bulge formation rate.

Stellar feedback seems to be crucial in determining whether or not clumps live long enough to have an important effect on their host galaxy. However, properly accounting for environmental effects such as mergers and gas accretion from large-scale filaments can be just as important, because these processes affect the overall state of the host galaxy. Isolated, gravitationally unstable disks are inherently artificial and those galaxies quickly exhaust their gas. Since there is no fresh supply of gas, there can be no new clumps formed after the initial clumpy phase. Hence, using simulations with feedback models as realistic as possible and in a full cosmological context is the most promising way to improve our understanding of the nature of giant star-forming clumps and their significance for galaxy evolution.

In this work, we use the results of the FIRE (Feedback in Realistic Environments) project, a suite of cosmological zoom-in simulations with explicit treatment of stellar feedback (Hopkins et al. 2014). We analyse the formation and evolution of giant clumps in a simulated massive gas-rich galaxy at $2.2 \geq z \geq 1.0$. Section 3.2 describes the simulation used in this work. In Section 3.3, we outline the procedure employed to identify clumps in snapshots. In Section 3.4, we present distributions of various physical properties of giant clumps, such as their mass, size, elongation, gas fraction, and stellar age. We investigate how the occurrence of giant clumps changes with redshift and overall properties of the host galaxy. We re-run our simulation with high time resolution (taking snapshots every $\sim$ 1 Myr) over six 50 Myr periods in order to follow the evolution of individual clumps. This allows us to directly measure clump lifetimes and compare them to stellar ages of clumps. We discuss the implications of our results and compare them to previous studies of clumpy galaxies in Section 3.5. Finally, we summarize the conclusions in Section 3.6. In Appendix 3.A, we compare the properties of clumps found in this simulation with clumps from a few additional cosmological simulations of galaxies at high redshift, run at higher resolution, but over a more limited redshift range.

3.2 The Simulation

3.2.1 The FIRE project

The simulation analysed in this study is part of the FIRE project[2], a suite of high-resolution cosmological zoom-in simulations of galaxy evolution. A detailed description of the numerical methods and physics implemented in the FIRE simulations can be found in Hopkins et al. (2014) and references therein. Here we briefly summarize the most important features.

The FIRE simulations were performed with the newly developed GIZMO code[3], using a pressure–entropy formulation of the smoothed particle hydrodynamics equations (P-SPH). P-SPH implementation gives good agreement with analytic solutions and grid codes on a broad range of test problems (Hopkins et al. 2013), thereby resolving some well-known issues of density-based SPH codes and their differences with grid-based codes (Agertz et al. 2007). The gravity solver used is a modified version of GADGET-3 (Springel 2005). The initial conditions for our simulations were generated with the MUSIC code (Hahn & Abel 2011) and are part of the AGORA comparison project (Kim et al. 2014).

The FIRE simulations include explicit treatment of the multi-phase ISM consisting of molecular, atomic, ionized and hot diffuse components. This was made possible by the high resolution and the improved modeling of gas cooling and heating. The resolution of the FIRE simulations is sufficiently high to resolve the turbulent Jeans mass/length of interstellar gas. The initial particle mass in the high resolution region of the simulation we analyse in detail is 3.7×10^5 $M_\odot$ for gas and star particles, and 2.3×10^6 $M_\odot$ for dark matter. Force softening lengths are adaptive, with the minimum fixed past $z \sim 10$ at 20 pc (physical units) for baryons and 210 pc for dark matter.

Star formation proceeds in regions which are (i) dense (above density threshold $n_{th} = 10$ cm^{-3}), (ii) molecular and (iii) self-gravitating. Gas particles which meet all of these criteria form stars at 100% efficiency per free-fall time. This is motivated by high-resolution simulations of turbulent media (e.g. Padoan & Nordlund 2011; Ballesteros-Paredes et al. 2011; Padoan, Haugbolle & Nordlund 2012) which suggest that small-scale star formation efficiency should be high; however, our results are not very sensitive to the exact value we chose. The galaxy-averaged

[2] http://fire.northwestern.edu/

[3] A public version of GIZMO is available at http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html.

efficiency of $\sim 1\%$ per dynamical time is regulated by stellar feedback and basically independent of the small-scale star formation law (Hopkins et al. 2011; Ostriker & Shetty 2011; Faucher-Giguère et al. 2013).

The FIRE simulations implement multiple modes of stellar feedback: stellar radiation pressure, photoionization and photoelectric heating, energy and momentum feedback from Types I and II supernovae, stellar winds from asymptotic giant branch and O-type stars. The input parameters for modeling feedback are *not* adjusted 'by hand' to produce results that match observations. They are taken directly from stellar population models in STARBURST99 (Leitherer et al. 1999), assuming a Kroupa 2002 initial-mass function for masses in the range $0.1 - 100\ M_\odot$. We do not consider feedback from active galactic nuclei (AGN).

Explicit treatment of the ISM and feedback physics allow the FIRE simulations to successfully reproduce many observed galaxy properties, such as the stellar mass–halo mass relation, Kennicutt–Schmidt law, and the star-forming main sequence (Hopkins et al. 2014). They show a good agreement with the observed mass–metallicity relations (Ma et al. 2016). Furthermore, our simulations predict neutral hydrogen covering fractions around high-redshift galaxies consistent with observations (Faucher-Giguère et al. 2015; Faucher-Giguère et al. 2016). The FIRE simulations also produce galactic-scale winds whose velocities and mass loading factors satisfy observational requirements (Muratov et al. 2015). These results give us confidence that our simulations can reliably reproduce star formation and ISM properties on galactic scales, which are crucial for investigating the nature of giant star-forming clumps and their impact on the host galaxy.

3.2.2 The `m13` simulation

In this work, we study in detail the most massive galaxy in the suite of FIRE simulations which is described in Hopkins et al. (2014), named `m13`. The name of the galaxy comes from the fact that it resides in a dark matter halo of mass $\sim 10^{13}\ M_\odot$ at $z = 0$, corresponding to a halo hosting a small galaxy group. We choose to focus on this galaxy because it is the only galaxy in the original FIRE sample which is as massive at $z \sim 2$ as most observed clumpy galaxies at this redshift – its stellar mass[4] at $z = 2$ is $\sim 4 \times 10^{10}\ M_\odot$, matching the stellar mass of clumpy galaxies in Genzel et al. (2011). The mass of `m13`'s dark matter halo at $z \sim 2$ is $8.7 \times 10^{11}\ M_\odot$.

[4] All global properties of the galaxy are computed in a 10-kpc sphere around the galaxy's center of mass.

We analyse the evolution of the m13 galaxy from $z = 2.2$ to $z = 1.0$. During this period, the galaxy changes from being a very clumpy, gas-rich system which forms stars vigorously to a more regular and moderately star-forming galaxy, similar to those at lower redshift. By studying the transition of one galaxy from its clumpy to non-clumpy phase, we wish to determine what causes this transformation and what long-term effects clumps might have on the evolution of their host galaxy.

In order to simulate a galaxy as massive as those observed at high redshift to $z = 0$, we adopted a resolution lower than in the rest of the original FIRE sample. In Appendix 3.A we compare the main properties of clumps in m13 with clumps found in a subset of simulations from the MassiveFIRE sample (Feldmann et al. 2016). This suite of simulations includes more massive galaxies than the original FIRE sample, run at 16 times higher resolution. However, they have only been evolved down to $z \sim 1.7$ due to their high computational cost. When possible, we also compare our results to the results of Hopkins et al. (2012), who analyzed the properties of star-forming clumps in high-resolution simulations of isolated galaxies implementing similar models of stellar feedback. We do caution that the stellar feedback implementation in that study was based on an older code, with a different hydro solver, and less accurate treatment of feedback processes compared to the updated simulations here (as discussed in Hopkins et al. (2014)). However this does not appear to change the qualitative conclusions therein.

During the analyzed period, the galaxy experiences interactions with a few smaller systems, but no major mergers. The most notable merger is that with a galaxy whose baryonic mass is ~ 20 times smaller than the baryonic mass of m13. The merger begins at $z \sim 2.3$ (that is ~ 150 Myr before the first snapshot analyzed in our series) and the secondary loses almost all of its gas during the first passage near the center of m13. What is left is a stellar component that is discernible in maps of stellar surface density all the way down to $z = 1$, as is orbits around m13.

We record snapshots of m13 and its surroundings every $\sim 15 - 40$ Myr, depending on the redshift. Thus we obtain a suite of 111 snapshots, which we refer to as the original series. We analyze these snapshots to determine the properties of clumps and how they change with redshift. In addition to that, we re-simulate six 50 Myr periods, starting at redshifts of 2.0, 1.9, 1.8, 1.7, 1.6, and 1.5, over which we follow the evolution of the m13 galaxy with fine time resolution, taking snapshots every $\sim 0.5 - 5$ Myr. This results in an additional 170 snapshots. The short time-scale between these snapshots allows us to track individual clumps and put constraints on

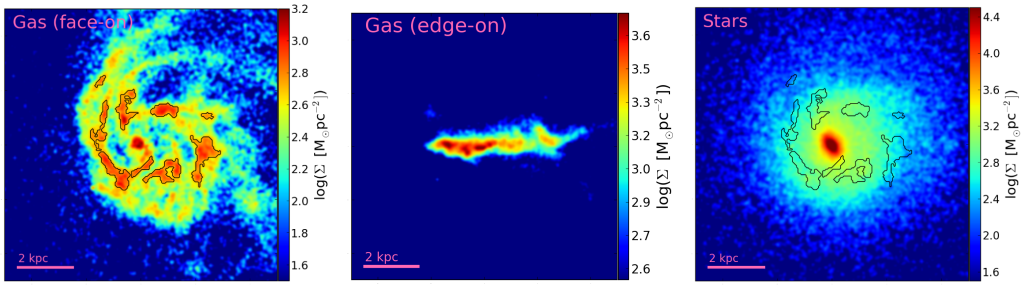

Figure 3.1: Surface density of gas (left and middle) and stars (right) for the m13 galaxy at $z = 2.0$. The size of the region shown is 10 kpc $\times$ 10 kpc (physical). Contours in the face-on projections show gas clumps identified by our clump-finding procedure (ignoring the central clump).

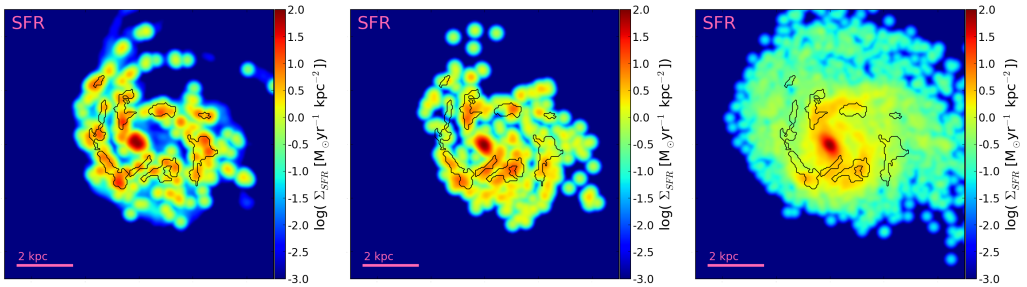

Figure 3.2: Surface density of the SFR calculated over three different time-scales (instantaneous on the left, averaged over 10 Myr in the middle, and averaged over 100 Myr in the right) for the m13 galaxy at $z = 2.0$. As in Figure 3.1, the size of the region shown is 10 kpc $\times$ 10 kpc (physical). Contours show gas clumps identified by our clump-finding procedure (ignoring the central clump).

their lifetime.

3.3 Clump Identification

To identify giant star-forming clumps, we assume that they correspond to regions of the disk with high surface density of gas. We choose this two-dimensional criterion for the definition of clumps in order to more easily compare our results with those from observations of real galaxies and mock images created from simulations. We tried using other criteria for the definition of clumps, ones which are more related to the physical (3D) density of gas and we find that they give very similar results when it comes to identifying the largest clumps. In observations of high-redshift galaxies, clumps are usually identified based on UV and optical light, serving as tracers of star formation. Here we use gas surface density to identify clumps, which is a good approximation because maps of the projected star formation rate (SFR) show that the

locations of gas clumps coincide with regions of intense star formation (see Figures 3.1 and 3.2). In general, we do not observe obvious clumps in the distribution of stars or dark matter coincident with the gas clumps. Discernible star clumps correspond to a few (in total, not per snapshot) smaller galaxies orbiting around `m13`.

As a first step in our clump-finding procedure, we find the total angular momentum of gas particles within 10 (physical) kpc from the center of mass of the galaxy and determine the Euler angles of the face-on projection in the simulation's coordinate system. Next, we form a two-dimensional 10 kpc $\times$ 10 kpc grid in the face-on projection, with cell size of 50 pc. We project the location of every particle within 10 kpc radius from the center of the galaxy on to the grid and for the sake of computational convenience deposit the entire particle into the corresponding cell (i.e. ignoring the particle's SPH kernel). We construct the gas surface density map by dividing the entire gas mass contained in a cell with its surface area. Using all cells with non-zero surface density, we find the mean value of gas surface density and the standard deviation. We use these two values in our definition of clumps– clumps correspond to regions which are at least one standard deviation above the mean surface density. The threshold values are chosen in this way because they provide the best match to what we visually identify as clumps in the gas surface density maps. Appendix 3.B contains more details on how changing the threshold value may affect our results. Finally, we smooth the gas surface density field using a Gaussian filter with the standard deviation of 50 pc (full width at half-maximum of $\sim$ 120 pc) and the radius of the Gaussian kernel of 450 pc. The $\sim$ 100 pc scale roughly coincides with the lower end of the reported size distribution of observed giant clumps at high redshift. This smoothing scale should be large enough to mitigate the effect of ignoring the SPH kernel in creating gas surface density maps.

To find clumps in the gas surface density map, we make use of a Python package ASTRODENDRO[5] to compute dendrograms – trees of hierarchical structure in our simulated data. To find clumps, we need to specify three parameters: (1) the minimum value of gas surface density which is used as a threshold to define clumps, which we set to be equal to one standard deviation above the mean surface density of all non-empty (unsmoothed) cells in that snapshot; (2) the minimum difference in the surface density between a large structure and any substructure within it, which we set to 10% of the mean surface density (this parameter is not very important in our particular study, since we primarily care about large clumps, and not so much

[5] `http://www.dendrograms.org/`

the substructures within them); (3) the minimum number of pixels within a clump which we set to 20, corresponding to a minimum effective radius of selected clumps of about 125 pc.

After computing a dendrogram for each snapshot using the described procedure, we obtain a list of clumps which satisfy the criteria we imposed. If a selected region has substructure, we decide what is going to be considered a 'clump' (whether the larger structure as a whole or smaller substructures individually) by visual inspection. In most cases we choose the largest structure to represent the clump and we follow the evolution of this object as a whole, without keeping track of individual substructures within it. However, in a few instances, the largest structure was branching out in such a way that it would be unlikely that it would be recognized as a single clump in real observation. In such cases, we selected the smaller, more compact substructures within that region to be identified as clumps. Many snapshots contain a clump located in the center of the galaxy, which corresponds to the bulge of the galaxy – we disregard those clumps from further analysis.

Using the ASTRODENDRO package to select clumps allows us to easily measure the size of each clump in the plane of the disk (xy plane), and express it in terms of its area (A), effective major (a) and minor (b) axes. The axes for each clump are calculated by ASTRODENDRO from the contour of the gas clump by fitting an ellipse to it and deriving the parameters of the ellipse. We calculate the effective radius (R) of a clump from its area (computed by ASTRODENDRO) as $R = \sqrt{A/\pi}$. We assume that the extent of each clump in the z-direction (perpendicular to the plane of the disk) equals $2R$ centered on the densest part of the clump. For most clumps, the enclosed mass does not heavily depend on the exact choice of its vertical extent, as long as it is roughly on the order of disk scale-height. Even if we extend clumps to 10 kpc in the z-direction (while keeping the surface area in the plane of the disk fixed), their mass increases only by a factor of a few (for the most massive clumps that factor is $\lesssim 2$)

All gas and stellar particles within the volume of space which is identified as a clump, belong to that clump. These selection criteria were devised to match the observational identification of clumps and do not tell us anything about whether clump particles are bound together or not. We do not subtract the underlying disk (stars or gas) from the clumps. Some observers subtract the background before evaluating the clump luminosity and mass (e.g. Förster Schreiber et al. 2011; Guo et al. 2012; Guo et al. 2015), while some do not (e.g. Elmegreen et al. 2009; Wuyts

et al. 2012). It is important to keep that in mind when comparing results from different groups, as the background subtraction may cause a factor of 2-3 difference in the estimated clump flux (Förster Schreiber et al. 2011; Guo et al. 2015).

Figure 3.1 shows the gas surface density of m13 at $z = 2$ in the face-on and edge-on projections (left-hand and middle panels) and the face-on distribution of stars (right-hand panel), which is fairly smooth and does not show prominent overdensities. The contours in the face-on projections mark the gas clumps identified in this snapshot. Figure 3.2 shows the surface density of the SFR calculated over three different time-scales (instantaneous, averaged over 10 Myr and over 100 Myr). For reference, the contours of gas clumps are also shown in the maps of the SFR. Most identified clumps (and the central clump, which is excluded from our analysis) correspond to regions of intense star formation and would likely be prominent in the rest-frame UV images and other tracers or recent star formation. However, there is no exact one-to-one correspondence between the gas clumps and the regions of elevated SFR, nor are the maps of the SFR calculated over different time-scales identical. This is in agreement with the results of Moody et al. (2014) who found that clumps identified in the gas maps are not necessarily found in the mock images or stellar maps of the same galaxy.

3.4 Results

3.4.1 Clump properties in the m13 simulation

The clump-finding procedure identified a total of 506 clumps in the original m13 series of snapshots at $2.2 \geq z \geq 1.0$ (not counting clumps located in the center of the galaxy, which we discarded). The average number of identified clumps per snapshot varies with redshift (as discussed in more detail in Section 3.4.2) from about eight clumps per snapshot at $z \sim 2$, to no clumps at all at $z \sim 1$. Figure 3.3 shows clump distributions in terms of their baryonic mass, physical size (effective radius), elongation (the ratio of the minor to major axis) and gas fraction (defined as $f_g = M_{gas}/[M_{gas} + M_{stars}]$). Even though the focus of this study is on giant clumps, which typically have masses in the range $10^8 - 10^9$ $M_\odot$, our clump-finding procedure is capable of finding clumps smaller than that, with masses $\sim 10^7$ $M_\odot$. The lower limit for clump mass is slightly different for every snapshot because it depends on the gas surface density threshold which varies from snapshot to snapshot. The highest surface density threshold over all snapshots is ~ 650 $M_\odot pc^{-2}$, which corresponds to a minimum clump mass of $10^{7.5}$ $M_\odot$. This is a conservative limit on our mass completeness over the entire analyzed redshift range. In many snapshots we are able

to identify clumps much smaller than that limit.

The number of clumps with total baryonic mass greater than 10^8 $M_\odot$ is 155 and their effective radii range from $\sim$ 150 to $\sim$ 600 pc, which is consistent with the properties of observed giant clumps at high redshift (e.g. Förster Schreiber et al. 2011). Clump gas fractions vary from highly gas-dominated clumps which have over 90% of their mass in gas, to clumps with $f_g \sim 0.3$. Most giant clumps have comparable mass contributions from gas and stars, with the median value for giant clumps of $f_g = 0.57$. Morphologies of giant clumps in our sample are very diverse, from nearly spherical clumps with the ratio of minor to major axis close to unity ($b/a \approx 1$) to highly elongated clumps with minor axis almost an order of magnitude smaller than the major ($b/a \sim 0.1$).

3.4.2 Occurrence of clumps over redshifts

The number of massive clumps present in the galaxy decreases with time over the redshift range $2.2 \geq z \geq 1.0$ which spans almost 3 Gyr. Figure 3.4 shows the mass distribution of clumps in three redshift bins. The number of identified clumps clearly decreases with decreasing redshift. The clumps also get smaller with time; the mean and median of the mass distribution, indicated by solid and dashed lines, respectively, move towards lower masses at lower redshifts.

If large clumps do form due to gravitational instabilities in gas-rich disks, then their occurrence rate might be correlated with the gas fraction of the galaxy. Figure 3.5, top panel, shows the galaxy's gas fraction and the gas velocity dispersion in the z-direction, both as functions of redshift. We calculate the gas fraction by taking into account gas in all phases. The line-of-sight velocity dispersion calculation also includes all gas particles–we do not exclude any inflows/outflows which may be present. The galaxy is very gas-rich at $z > 2$, with gas fraction above 30%, which is consistent with observations of high-redshift galaxies (Tacconi et al. 2008; Tacconi et al. 2010). With time, the gas fraction decreases, settling at about 10 – 15% around $z \sim 1$, when the stellar mass of the galaxy is $\sim 6 \times 10^{10}$ $M_\odot$. The gas velocity dispersion follows the same trend, though in a more bursty manner. The middle panel shows the SFR in the galaxy as a whole and in clumps. Clumps are sites of intense star formation which account for 10 – 60% of the total SFR in the galaxy at any given time, even though they make up only a few percent of its total mass. About 10^{10} $M_\odot$ of stars is formed in clumps during the clumpy phase in the evolution of the galaxy ($2.2 \gtrsim z \gtrsim 1.3$), which is ~20% of the total star formation

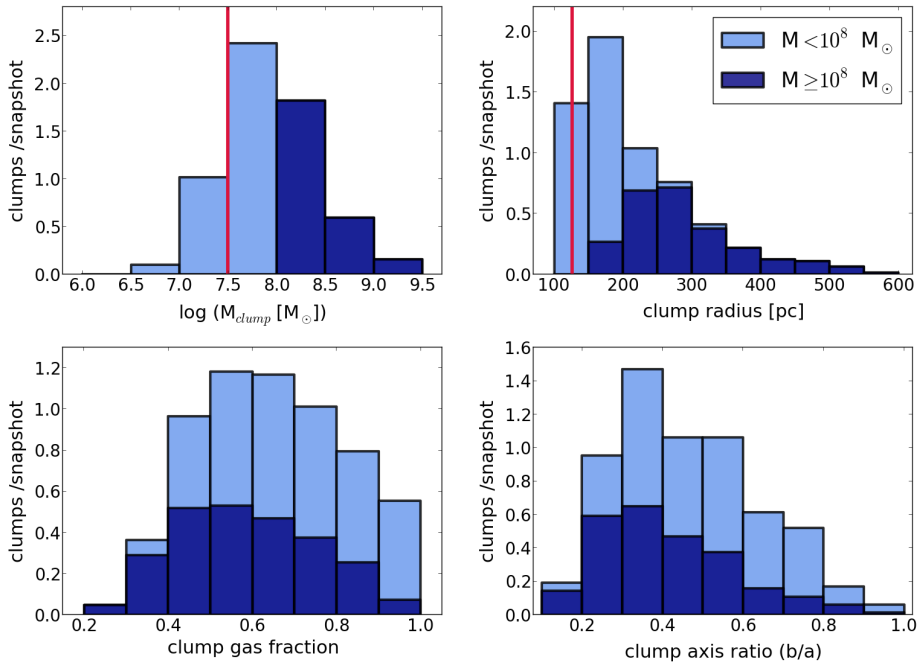

Figure 3.3: Distribution of clump properties in the `m13` simulation from $z = 2.2$ to $z = 1.0$. The upper-left panel shows the distribution of the total clump baryonic mass. The red line indicates the mass limit above which our clump sample is complete. Clump gas fraction, obtained as the ratio of clump gas mass to the sum of gas and stellar mass is shown in the lower left panel. The upper-right panel shows the distribution of clump effective radius, calculated from clump's surface area (A) as $R = \sqrt{A/\pi}$. The red line marks $R = 125$ pc, the lower limit on the radius, set by our requirement that clumps span the area of at least 20 pixels of size 50 pc $\times$ 50 pc. The distribution of the ratio of minor (b) to major (a) axes is shown in the lower right panel. Dark blue histograms represent giant clumps with baryonic mass $\geq 10^8$ M$_\odot$, and light blue is for less massive clumps (the light blue bars lie on top of the dark ones).

over that period. The SFR experiences a similar decrease with redshift as the gas fraction. The bottom panel shows the mass of individual giant clumps with mass $\geqslant 10^8$ M$_\odot$ as a function of redshift. The most massive clumps approximately follow the expected scaling with the gas fraction given in equation (3.3)($M_{cl} \sim f_g^3 M_{gal}$). There are no clumps with mass above 10^8 M$_\odot$ at $z < 1.3$, when the gas fraction falls under $\sim 15\%$. At these redshifts, the clump masses are in the range $10^6 - 10^8$ M$_\odot$, i.e. comparable to the mass of 'normal' GMCs found in the Milky Way. The properties of clouds in this mass range, found in simulations of isolated galaxies which implement a similar model of stellar feedback as our simulation, are analyzed in detail in Hopkins et al. (2012).

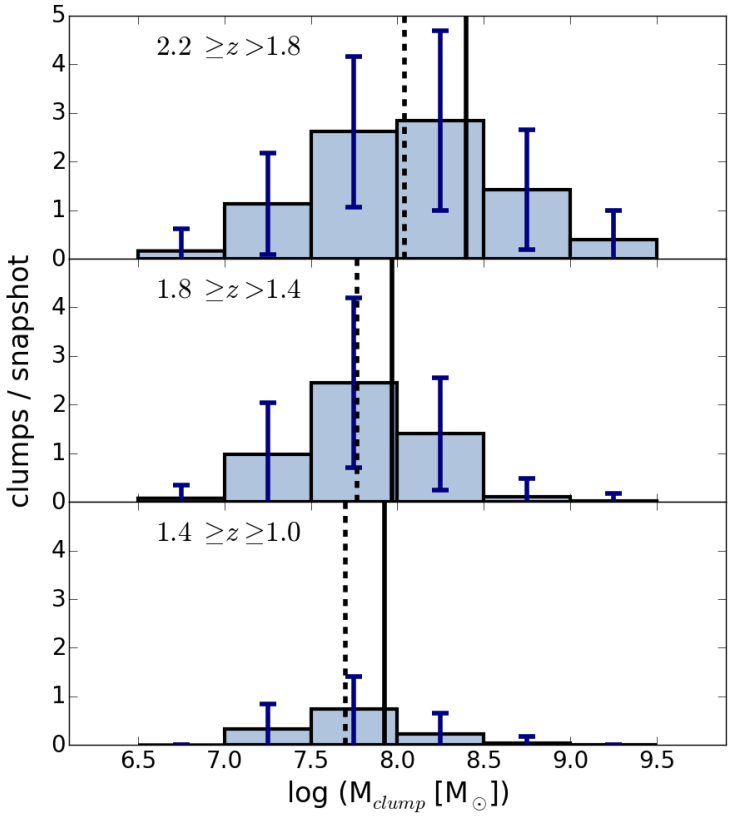

Figure 3.4: Distribution of clump baryonic mass in three redshift bins. Histograms indicate the average number of clumps per snapshot in each bin, whereas the error bars mark the standard deviation. The number of clumps at low redshift is reduced compared to high redshift and the clumps are less massive, as indicated by the mean and median of the distribution (represented by solid and dashed lines, respectively) which move over to lower mass values at lower redshift.

There is a relatively short period around $z \sim 1.7$ when the gas fraction slightly increases due to a minor merger with a galaxy of stellar mass $\sim 10^9$ M$_\odot$. The gray band in Figure 3.5 marks the period during which the small infalling galaxy is within 30 kpc from the center of mass of the whole system.

3.4.3 Stellar age in clumps

Figure 3.6 shows the mass-weighted age of star particles in giant clumps, as a function of the clump's galactocentric radius, i.e. the distance between the center of the galaxy and the clump's center of mass, color-coded by redshift. Mass-weighted stellar ages of clumps are typically a few hundred Myr; however, they span a broad range from tens of Myr for a few of the smallest clumps to about a

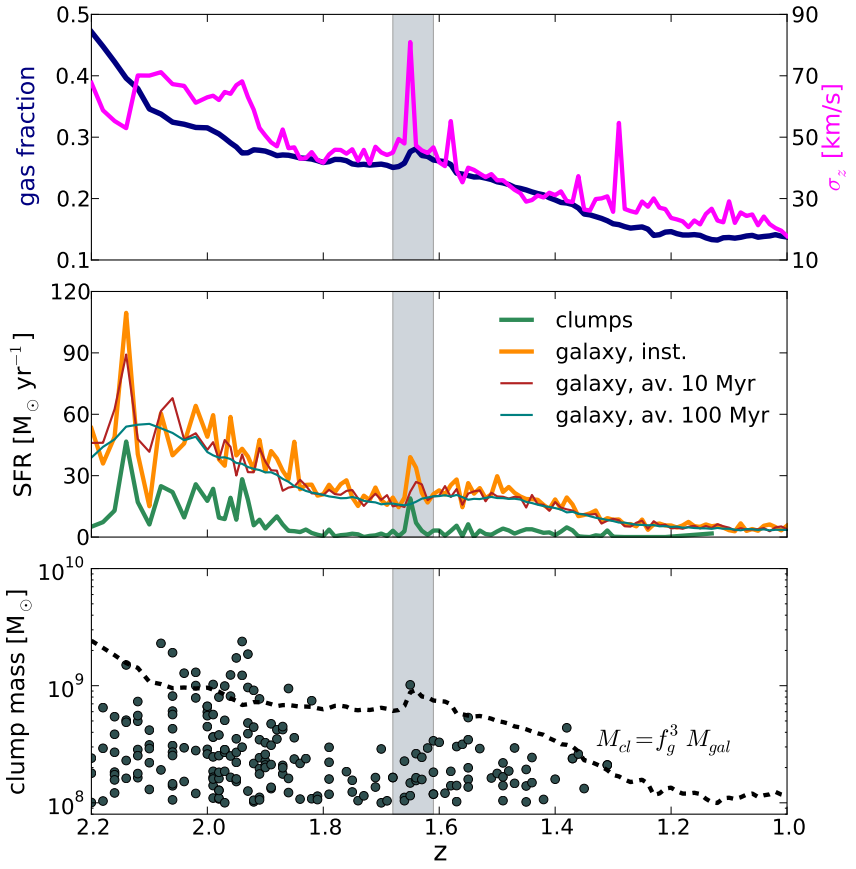

Figure 3.5: Upper panel: blue line represents the gas fraction of the m13 galaxy as a function of redshift. Gas velocity dispersion perpendicular to the plane of the disk is represented by the violet line. Middle panel: SFR in the whole galaxy (orange for instantaneous, red and teal for the SFR calculated by averaging over 10 Myr and 100 Myr time-scales, respectively) and in clumps (green) show that 10 – 60% of the total star formation occurs in clumps at any given time. Averaged over the analyzed time period, clumps account for $\sim$ 20% of the total star formation in the galaxy. Bottom panel: mass of individual giant clumps (M$\geqslant 10^8$ M$_\odot$) at their corresponding redshift. The most massive clumps follow the expected scaling with the gas fraction and mass of the galaxy, $M_{cl} \sim f_g^3 M_{gal}$, represented by the gray dashed line. The velocity dispersion, SFR and the occurrence of giant clumps are all high when the gas fraction is high, and decrease with decreasing redshift. The gray band marks the period during which the galaxy experiences a close encounter with a smaller galaxy, resulting in an increase of the gas fraction, SFR, and number of giant clumps.

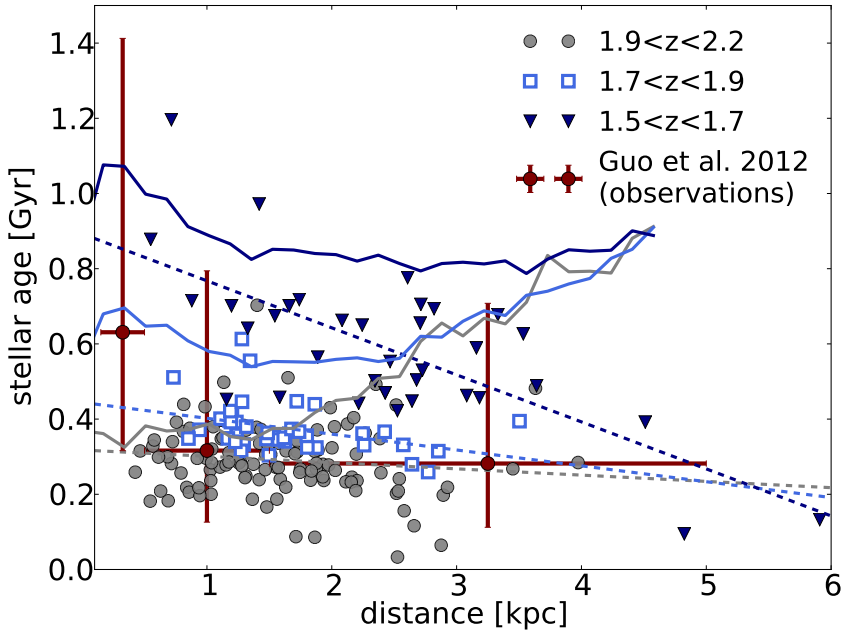

Figure 3.6: Mass-weighted stellar age in individual giant clumps with M$\geqslant 10^8$ M$_\odot$ as a function of distance from the center of the galaxy, color-coded by redshift. For all three redshift bins, we show linear fits (dashed) through the data points which indicate that clumps with older stellar populations tend to be located closer to the center of the galaxy. Average stellar ages of all the stars in the disk (including those in clumps), at the central redshift for each bin, are shown as solid lines. Clumps have slightly younger stellar populations than the disk average because they are regions of intense star formation and hence contain more young stars than the rest of the disk. Observational results from Guo et al. (2012) are show for comparison. We multiplied their x-axis values by the radius of *our* galaxy (R = 5 kpc).

Gyr. These numbers are consistent with the estimated ages of stellar populations in observed high-redshift clumps, which range from $\sim$ 10 Myr to several hundred Myr (Elmegreen et al. 2009; Förster Schreiber et al. 2011; Guo et al. 2012; Wuyts et al. 2012; Adamo et al. 2013). However, as we show below, stellar ages are not necessarily indicative of clump lifetimes.

3.4.4 Radial gradients in clump properties

Radial gradients in various clump properties, most notably in the stellar age of clumps, have been observed and interpreted as evidence of radial migration of clumps and their long survival. In this section, we show that our clumps, although short-lived, can reproduce similar trends with galactocentric radius, thus suggesting

that these observational results may not be as informative about clump survival as previously believed.

Most star-forming clumps in m13 are centrally located, within $\sim$ 3 kpc from the center, especially at high redshift. For a given redshift bin, there is a tendency for clumps with older stellar populations to be located closer to the center of the galaxy, as indicated by dashed lines in Figure 3.6 which represent linear fits through data points. This trend is only marginal in the high-redshift bin, but becomes more pronounced at lower redshifts. Anti-correlation of stellar age and distance from the center (shown in Figure 3.6) has been observed in high-redshift clumps by Förster Schreiber et al. (2011), Guo et al. (2012), and Adamo et al. (2013). Because clumps are regions of intense star formation, most of them contain slightly younger stellar populations than the disk average at the corresponding redshift (indicated by solid lines in Figure 3.6), which includes stars in clumps and inter-clump regions. The average stellar age increases with radius in the outskirts of the disk (i.e. beyond $\sim$ 3 kpc from the center). This is similar to the results of El-Badry et al. (2016), who analyzed stellar migration due to gas outflow/inflow cycles caused by stellar feedback in the FIRE simulations of low-mass galaxies. They found that most stars form in the central parts of the host galaxy and then exhibit systematic outward migration over the course of their lifetime, causing a positive stellar age gradient.

In Figure 3.7 we show the radial dependence of the clump stellar surface density, the specific star-formation rate (sSFR, i.e. the instantaneous SFR of a clump divided by its stellar mass), and gas fraction. We show data for all our clumps, color-coded by redshift as in Figure 3.6. We compare our results to observations by Guo et al. (2012) for the stellar surface density and the sSFR (we use their background-uncorrected results), and to the results from simulations by Mandelker et al. (2014) for the clump gas fraction. They report their results in terms of the fractional galactic radius (d/R, where R is the radius of the galaxy), hence we multiply their x-axis values by $R = 5$ kpc, which is the estimated radius of our galaxy. We bin our data in the same way as they did in order to make direct comparison easier. This makes our bins unevenly populated; while most of them have plenty of clumps, some have only a few members and should be regarded with caution.

Our results for the sSFR of clumps agree very well with the observations by Guo et al. (2012) and indicate that the sSFR of clumps slightly increases towards the outer regions of the disk. The scatter is large in both of our data sets, spanning almost the same range (from log(sSFR)$\lesssim$ -1 to log(sSFR)$\gtrsim$ 2). The mean values in

each bin are within one standard deviation from each other (with one caveat: our innermost bin contains only two clumps).

The middle panel of Figure 3.7 shows the radial dependence of the stellar surface density of clumps. The agreement with the results of Guo et al. (2012) is again fairly good.

Finally, we compare our radial dependence of the clump gas fraction with the results from simulations of Mandelker et al. (2014) which produced long-lived clumps. Although the radial trends are similar, our clumps have systematically higher gas fractions by ∼1 dex. Measuring the absolute gas fractions in high-redshift clumps may be more informative than measuring the radial gradients for the purpose of distinguishing between different (i.e. short- versus long-lived) clump models. This is consistent with the results of Mandelker et al. (2016) who found significantly higher gas fractions in the population of their short-lived clumps, compared to the gas fractions of their long-lived clumps (although the radial dependence of the gas fraction they obtained for the short-lived clumps is flatter than ours).

3.4.5 Clump lifetime

Snapshots used in the previous section to describe physical properties of clumps in `m13` were not useful for estimating their lifetimes. Typical time differences between two consecutive snapshots in the series are 15 − 40 Myr (it varies with redshift), which is too long to follow individual clumps from snapshot to snapshot as the galaxy's morphology can change considerably over that time-scale. In order to track individual clumps we re-run our simulation from six starting points and evolve it for 50 Myr in each case, taking snapshots every few Myr[6].

After identifying gas clumps in all additional snapshots using the same procedure as described in Section 3.3, we determine the history of each clump from snapshot to snapshot by visual inspection. We do not trace the evolution of clumps by following their member-particles (by their IDs) because we want to avoid making any assumptions regarding how the content of individual clumps may change in time.

Using this method we are able to constrain the lifetime of each clump–the time difference between the first and last snapshot in which that region of the disk is

[6]For re-runs starting at $z = 2.0$ and 1.5, we tried out several different time-steps (Δt) ranging from 0.5 to 5 Myr to find the most appropriate one–short enough to allow us to track individual clumps, but long enough to reduce the number of needed snapshots to a manageable size. We find that $\Delta t = 2$ Myr is a good compromise, and we use that time-step for the remaining four re-runs.

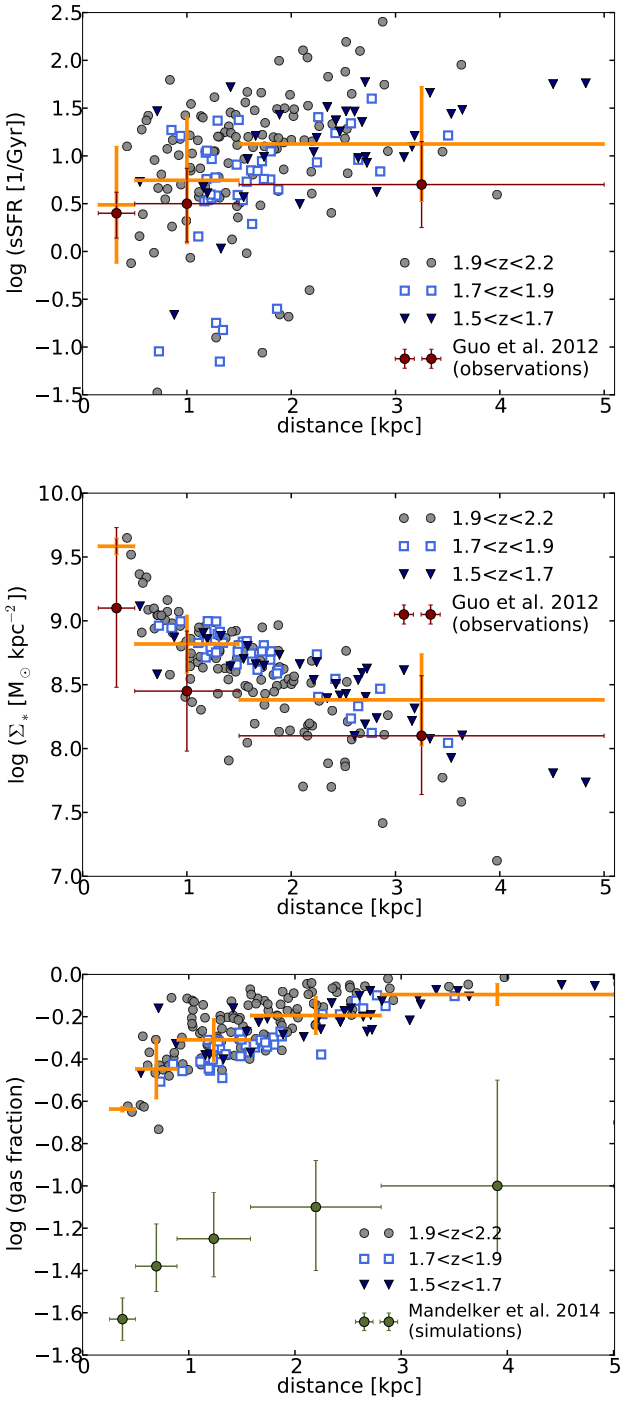

Figure 3.7: Different clump properties shown as functions of the clump's distance from the galactic center. The top and middle panels show the sSFR and the stellar surface density of our clumps, along with the observational results of Guo et al. (2012). The bottom panel shows the clump gas fractions, compared to simulations of Mandelker et al. (2014). The mean values and standard deviations of our data, binned in the same way as the data we compare them with, are shown in orange.

identified as a clump. For clumps which are already present in the first or still present in the last snapshot in a 50 Myr series, we are only able to set a lower limit on the lifetime using this method. For that reason, we select a sample of clumps–those which are *formed within the first 10 Myr* of a re-run–and use only those clumps for the lifetime analysis. In other words, we exclude all clumps which are already present in the first snapshot in a 50 Myr series (because we do not know when these clumps were formed) and all clumps which form later than 10 Myr into a re-run (because they might not have enough time in the remaining period to fully evolve).

Figure 3.8 shows the lifetime of each clump in this sample as a function of its maximum mass. Lifetimes range between 2 and 48 Myr, and the mean lifetime of clumps with mass above 10^8 $M_\odot$ is $\sim$ 22 Myr. There were no clumps which formed within the first 10 Myr of a 50 Myr re-simulated period which managed to survive until the very end–they all got disrupted before the last snapshot in that series (except for one clump, marked by a yellow square in Figure 3.8, which sank to the center of the galaxy and merged with the central overdensity, which is excluded from our clump sample by definition). Smaller clumps sometimes end their lives by merging with more massive ones. If two or more clumps merge, the newly formed clump carries the designation of the most massive progenitor, while the smaller ones cease to exist. More massive clumps typically live longer than less massive ones, which is consistent with the results from isolated disk simulations in Hopkins et al. (2012). In Section 3.5.2, we discuss the fate of stars formed inside of a clump after the gas gets disrupted.

3.5 Discussion

3.5.1 Clump formation and destruction

Clump formation in gas-rich systems

We follow the evolution of a massive, gas-rich, star-forming galaxy from $z = 2.2$ to 1.0 and investigate the nature of massive gas clumps formed within it. During the first $\sim$ 1 Gyr ($2.2 \gtrsim z \gtrsim 1.5$), when the gas fraction and velocity dispersion are both high (see Figure 3.5), the morphology of the gaseous disk is highly irregular and asymmetric, dominated by a few clumps of increased gas surface density. Some of the clumps look compact and nearly spherical, whereas some are elongated filaments which often break up into smaller sub-clumps. At later times ($z \lesssim 1.5$), as the gas fraction and velocity dispersion drop to roughly half of their initial value and the

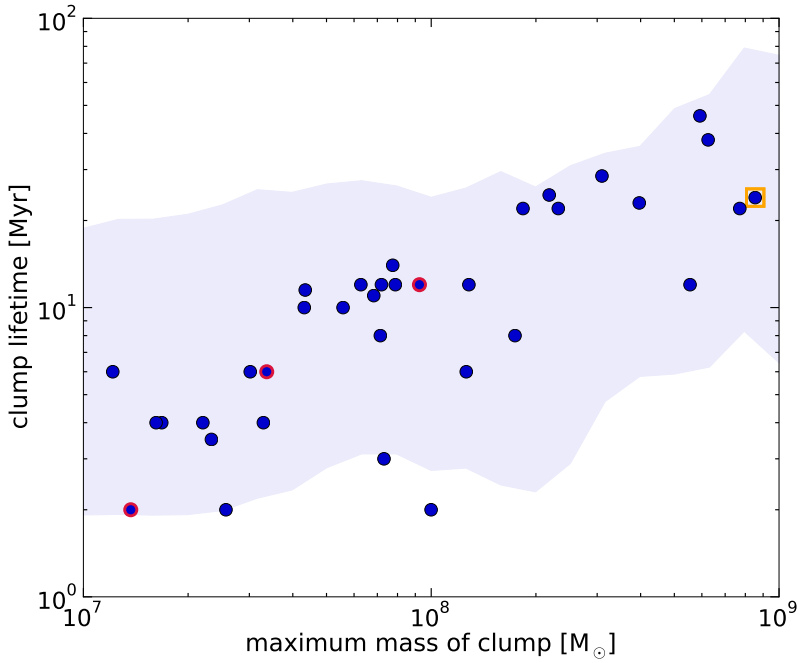

Figure 3.8: Clump lifetime as a function of clump's maximum mass. We show only clumps which are formed within the first 10 Myr of a re-simulated period. Clumps which end their lives by merging with other clumps are indicated in red. The clump marked by a yellow square ended its life by falling to the center of the galaxy. The shaded region shows the 10 – 90% range of clump lifetimes as a function of their mass, for clumps found in high-resolution simulations of isolated galaxies by Hopkins et al. (2012).

cold gas component stabilizes into a more regular disk of lower surface density, the gas overdensities (i.e. clumps) become smaller and less frequent. This change with redshift is related to the lack of gas-rich mergers and decreased gas accretion. At later times there are no obvious filaments of in-flowing gas and the overall gas content of the galaxy goes down.

Stellar clumps

The analyzed clumps are identified in the gas surface density maps of the `m13` galaxy. We do not see obvious clumps in stellar maps which include all stars (young and old). If we look at the young stellar population only, we see a certain level of clumpiness, as shown in the maps of the SFR (Figure 3.2). Prominent substructures in the total stellar component are associated with smaller galaxies falling into the potential well of `m13`, orbiting around and merging with it. One merger event, with a galaxy of

initial stellar mass of 1.6×10^9 $M_\odot$, starts a few tens of Myr before our first analyzed snapshot, and lasts until $z \sim 1$. Another galaxy with stellar mass $\sim 10^9$ $M_\odot$ passes through the m13's potential well around $z \sim 1.6$. Both of these smaller galaxies lose or exhaust their gas during the first passage through the densest regions of m13. After that, they appear only as stellar overdensities and we do not identify them as (gas) clumps. However, their presence might contribute to the disturbed state of the gaseous disk of m13.

Clump destruction and lifetimes

Giant gas clumps in m13 have relatively short lives, of the order of 10^7 yr. There is a correlation of clump lifetime and mass – more massive clumps tend to live longer. Smaller clumps can end their lives by merging with more massive ones or by dissolving as their surface density drops below the threshold which is used to define clumps. Massive clumps often end their lives by breaking up into smaller pieces, which can form a ring-like structure, indicating that the clump was disrupted from within by its own stellar feedback. These fragments often merge with other gas clouds to form new clumps.

The lifetimes of clumps which we measure in m13 are consistent with the result of Hopkins et al. (2012), as shown in Figure 3.8. They studied clumps in isolated galaxies with similar implementation of stellar feedback as is used in this work but at much higher resolution. Even though their methods for finding clumps and measuring clump lifetime are different from those which we employ, the conclusions derived from their work and ours are very similar. On the other hand, cosmological simulations which include only thermal feedback from supernovae and stellar winds, (Ceverino et al. 2010; Ceverino et al. 2012; Mandelker et al. 2014) often produce giant clumps which live long lives ($\gtrsim 10^8$ yr) and migrate to the center of their disks. Unlike the FIRE simulations, many previous studies do not resolve the evolution of supernovae and hence do not account for the momentum accumulated in the process. Direct comparison between our results and the results presented by these authors is complicated due to many differences in the way we define clumps and measure clump properties; however, it seems that including radiation pressure feedback in their simulations (although their treatment of radiation pressure is different from ours) decreases the lifetime of their giant clumps and makes our results qualitatively more similar (Moody et al. 2014; Mandelker et al. 2016).

Another difference between our simulation and some of the previous analyses of

clumpy galaxies, especially those using non-cosmological simulations, is our moderate value of the galaxy gas fraction ($f_g \sim 0.3$ at $z = 2$, roughly 70% of which is in the cold phase characterized by temperature $\lesssim 10^4$ K) compared to values of the molecular gas fractions observed in high-redshift galaxies which reach as high as ~0.4 - 0.6 (Tacconi et al. 2008; Tacconi et al. 2010). In their studies, Dekel & Krumholz (2013), Mandelker et al. (2014), and Bournaud et al. (2014) argue that clumps can accrete gas from the surrounding gas-rich regions, thus prolonging their lives by replenishing the gas they lost due to stellar feedback and star formation. Since we do see a correlation between high gas fraction and the occurrence of massive clumps, and also between the mass of clumps and their lifetime, it is possible that having higher gas fractions in our simulation might produce some more massive clumps with longer lifetimes. However, it is unlikely that this effect alone can explain the discrepancy in estimates of clump lifetimes, because the results of Hopkins et al. (2012), who studied isolated disks with high gas fractions (initial $f_g \sim 0.6$), found clump lifetimes not too different from ours using a similar implementation of stellar feedback.

Different definitions of clumps may also lead to differences in the estimates of clump lifetimes. Increasing the density threshold (or the minimum clump surface area) would lead to shorter clump lifetimes because any region of the disk would be less likely to meet these more restrictive criteria in order to be identified as a clump. If we used a more restrictive definition of clumps which identified only very dense structures (more similar to what we call sub-clumps) as clumps, it is likely that this would result in even shorter clump lifetimes than those shown in Figure 3.8.

Comparing the clump lifetime (Figure 3.8) with the age of stellar population in clumps (Figure 3.6), we find that the latter is about an order of magnitude greater than the former. This indicates that a significant fraction of stars which reside in clumps were not actually formed there, but belong to an older stellar population whose location happens to coincide with the clump or which have been accreted by the clump. This 'contamination' by older stars may be more significant in the inner regions of the galaxy, which are more densely populated than the outskirts. This may explain the observed anti-correlation between galactocentric radius and stellar age of clumps. A similar anti-correlation exists in the overall stellar population in the inner parts of `m13` and could simply be translated to clumps. Therefore, measuring stellar ages of clumps might not be a reliable way to establish the absolute age of clumps themselves or the relative age of one clump to another.

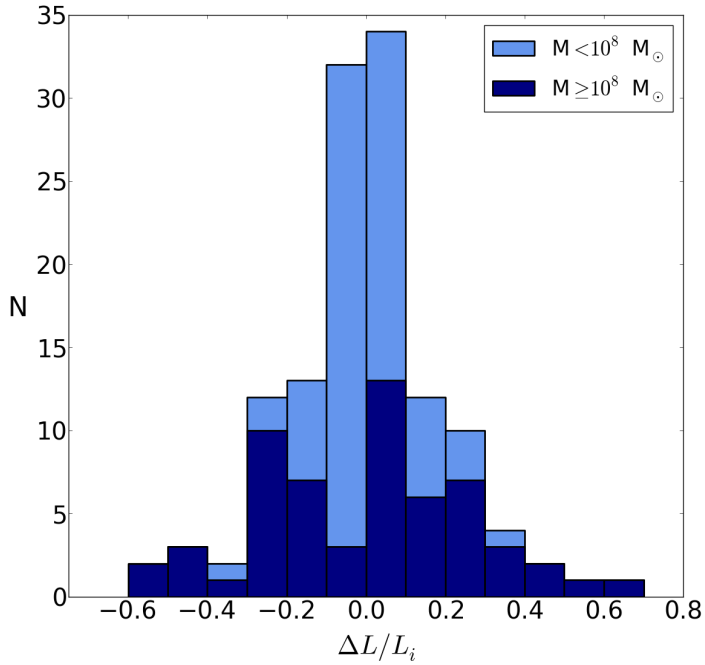

Figure 3.9: Fractional change of clump's orbital angular momentum per unit mass over the course of clump lifetime. Clumps change their specific orbital angular momentum by up to $\sim 50\%$, and they are equally likely to gain angular momentum as they are to lose it. This result does not support the scenario in which clumps systematically migrate towards the center of the galaxy.

Clump orbits and (lack of) inspiral

On average, clumps follow the overall rotation of the disk, with some deviations, presumably induced by gravitational torques. Several clumps end up reaching the center of the galaxy, but interestingly most of those do start their descent from nearly the same location in the disk. Some other clumps at similar distances from the center, but at different locations, have very different trajectories and do not end up in the center.

The change in the clump's galactocentric radius is related to its loss of orbital angular momentum. To get a sense of how this quantity changes over the lifetime of a clump, in Figure 3.9 we show the difference in the specific orbital angular momentum of clumps in the z-direction, i.e. the sum of angular momenta (with respect to the center of the galaxy) of all clump particles (gas and stars) divided by the mass of

the clump,

$$L = \frac{\sum_j m_j \boldsymbol{r}_j \times \boldsymbol{v}_j}{\sum_j m_j} \cdot \hat{z} \,, \tag{3.4}$$

between the final and initial snapshot of that clump, expressed as a percentage of its initial angular momentum ($\Delta L/L_i = (L_f - L_i)/L_i$). We find that clumps experience a change in their orbital angular momentum by a few tens of percent (up to $\sim 50\%$) and that roughly equal numbers of clumps experience inward migration as outwards. Small clumps on average experience smaller changes in their angular momentum compared to more massive ones, due to their shorter lifetimes.

Our finding that giant clumps are equally likely to gain angular momentum over the course of their lifetime as they are to lose it indicates that clump migration on such short time-scales might be dominated by gravitational torquing and tidal forces due to the asymmetric state of the disk, inflowing gas streams and orbiting companion galaxies, rather than due to dynamical friction which acts on longer time-scales and would produce a 'one-sided' systematic trend of angular momentum loss, as seen in simulations which produce long-lived clumps (e.g. Ceverino et al. 2010; Ceverino et al. 2012; Bournaud et al. 2014; Mandelker et al. 2014).

3.5.2 Are clumps gravitationally bound?

Empirically determined values of Toomre Q parameter in clumpy galaxies at high redshift (Genzel et al. 2011) suggest that giant clumps reside in unstable regions of the disk, characterized by Q-values below unity. In order to compare our simulation with these observations, we create maps of the 'observed' Q parameter (for gas) by locally evaluating equation (3.1) using the line-of-sight velocity dispersion (σ_z) and the orbital frequency Ω (calculated from the local azimuthal velocity) instead of the epicyclic frequency κ.

Maps of Toomre Q values throughout the disk (Figure 3.10) indicate that the locations of many clumps coincide with regions of the disk which are gravitationally unstable ($Q < 1$). However, there is large scatter. Often, regions of low Q correspond to areas which are still in the process of turning into clumps. Similarly, 'mature' clumps which have already sustained substantial star formation can have values of $Q > 1$. This is not surprising because the Q parameter was developed in the context of linear perturbation theory and is only useful as a diagnostic of the onset of gravitational instability and the early phases of fragmentation. When the fragmentation is already under way, we move into non-linear regime and Toomre Q is no longer a valid diagnostic.

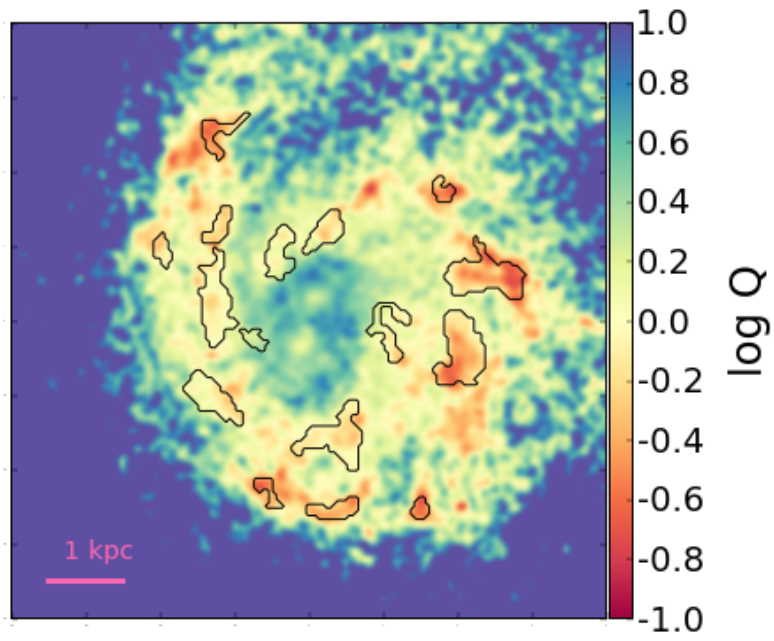

Figure 3.10: Map of Toomre Q parameter of the gas disk at $z \approx 1.9$. Clumps identified in this snapshot are overlaid as black contours. Most clumps, especially those in the early phase of their life, coincide with regions of low Q.

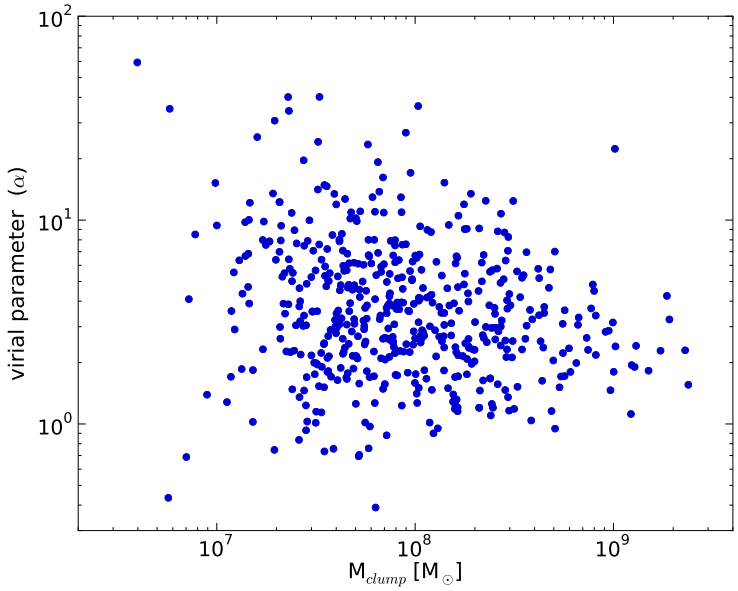

Figure 3.11: Virial parameter α of each clump in the original series of `m13` clumps as a function of their mass. The virial parameter is a ratio of clump's virial mass to its actual mass and can be used to evaluate whether a clump is gravitationally bound or not. Values of $\alpha \gtrsim 1$ indicate that clumps are only marginally bound or unbound.

Our results are in agreement with those of Inoue et al. (2016) who performed a detailed study of Toomre Q in simulated high-redshift clumpy galaxies. They found that regions of the disk where clumps form are typically characterized by lower values of Q compared to the interclump regions. However, their clumps often form in regions of $Q \geq 2-3$, indicating that the critical value of $Q = 1$ may not necessarily be a good criterion for fragmentation in realistic disks in which many assumptions of the conventional Toomre analysis break down. Their results may also be interpreted as an indication that some mechanism other than the Toomre instability may play a role in the formation of clumps in high-redshift disks.

Some previous studies of simulated clumpy disks (e.g. Ceverino et al. 2010; Bournaud et al. 2014; Mandelker et al. 2014) found clumps to be roughly spherical, gravitationally bound structures able to maintain their compact morphology for hundreds of millions of years. Clumps which we find in `m13` are qualitatively very different. Their morphology is more complex: some clumps seem compact, while others are very elongated, with prominent substructures along the filaments. The shape of clumps can change over the course of their (short) lifetime as clumps stretch, break up, and merge. This behavior does not support the picture of clumps as long-lived, virialized (gravitationally bound and relaxed) structures. The most probable reason for this qualitative difference between our clumps and clumps found in some of the earlier studies is the treatment of stellar feedback. Many simulations so far have used relatively simple models of feedback which mostly include thermal energy from supernovae, but ignore some other types of feedback present in the early phase of stellar evolution. The additional feedback modes which are included in the FIRE simulations, especially radiation pressure, may play an important role in determining clump properties, as discussed in Hopkins et al. (2012).

Figure 3.11 shows the virial parameter (α) of each clump in the original series of `m13` snapshots. It is defined as the ratio of the virial mass of each clump to its actual mass (Bertoldi & McKee 1992):

$$\alpha = \frac{M_{vir}}{M} = \frac{5\sigma^2 R}{GM} \,, \tag{3.5}$$

where σ is the 1D velocity dispersion (in our case σ_z) of gas particles in the clump and R is the clump radius. The virial parameter of a clump is related to the ratio of its kinetic to potential energy and can be used to gauge whether a clump will collapse under its own gravity. Structures with $\alpha \lesssim 1$ are considered gravitationally bound[7], whereas those with $\alpha > 1$ are unbound – they have enough kinetic energy

[7]The expression for the virial parameter (equation 3.5) can include a constant factor of order unity

to overcome its self-gravity and expand into the surrounding medium. The results in Fig. 3.11 tell us that some clumps might be gravitationally bound. However, a large fraction of them seem not to be, or at least not entirely. Smaller substructures within clumps must be self-gravitating, otherwise they would not be forming stars (that is one of our imposed criteria for star formation; see Section 3.2 for more details). Giant clumps may hence be regarded as massive, fragmented clump complexes, which could explain why they are easily sheared apart and destroyed by feedback.

Giant clumps, especially the most massive ones at $z \sim 2$, almost always show signs of substructure–from when they were formed until they get destroyed and broken up into smaller fragments. The remaining fragments often get 'recycled' into clumps by coming together or by being accreted on to a larger clump (clumps which end their lives by merging with larger clumps are highlighted in Figure 3.8). Similarly, Behrendt et al. (2016) find evidence of substructure within giant clumps in their high-resolution simulation of an isolated galaxy. They suggest that giant clumps seen at high redshifts might just be groups or clusters of smaller ($\sim$ 100 pc) clumps observed with coarse angular resolution. This is in agreement with the results of Tamburello et al. (2015) who also studied the properties of clumps in high-resolution simulations of isolated disks. They defined clumps as gravitationally bound objects and found that most clumps in their sample have mass of $\sim 10^7$ $M_\odot$, and that clumps with mass $\geq 10^8$ $M_\odot$ (the mass typically associated with giant clumps) are rare. Because Tamburello et al. (2015) use a different definition of clumps compared to our treatment (i.e. we do not assume the gravitationally bound state of clumps a priori), their clump sample might better correspond to what we call sub-clumps in this study.

Where do the clump stars go?

Clump lifetime tells us how long the clump exists as an overdensity of gas, but it does not give us insight into what happens to the stars which were formed in the clump after the gas has dispersed. To investigate that, we identify particles in our simulation which were part of a clump as gas particles and then turned into star particles during the lifetime of that clump (particle IDs are preserved as particles turn from gas to stars). Note that each star particle represents a star cluster of mass $\sim 10^5$ $M_\odot$. We follow the positions of these particles during and after clump's lifetime. We find that stars formed in the same clump do not remain close to each

which depends on the clump density profile. This may change the criterion for virial equilibrium ($\alpha \simeq 1$) by a factor of a few, as discussed by Mandelker et al. (2016).

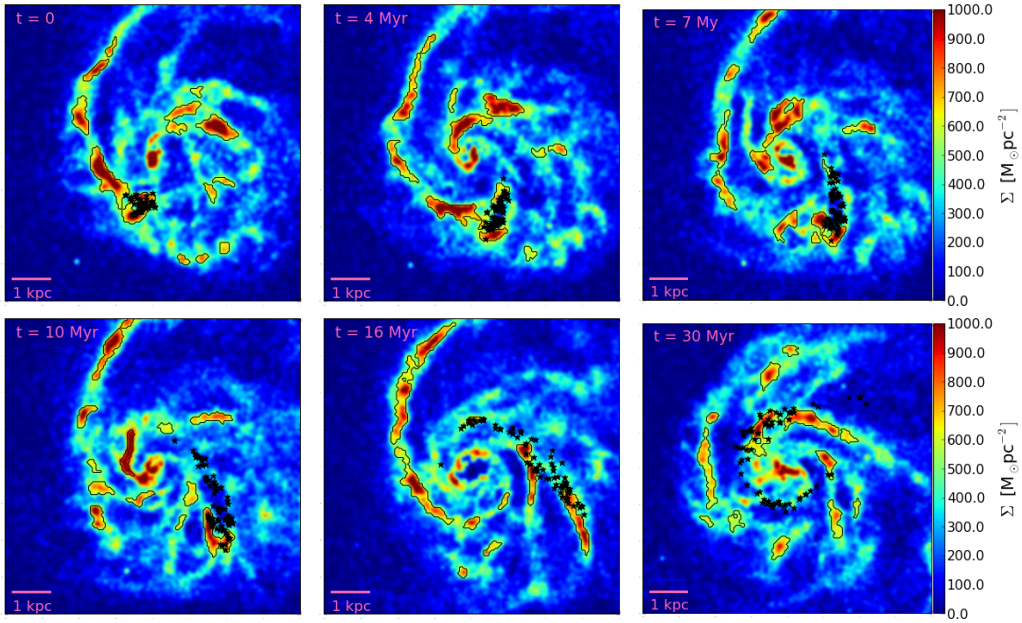

Figure 3.12: A series of snapshots spanning 30 Myr around $z \sim 1.8$. The colormap shows the gas surface density and the black contours represent gas clumps. Black asterisks mark stars which were formed in one of the clumps. The initial snapshot (upper-left panel) is taken at the time when the clump starts to break apart. At that point, all the stars are still located in the clump which formed them. Consecutive snapshots were taken at times indicated in the upper left corner with respect to the initial snapshot. Stars which were formed in the same clump are not bound; they gradually drift apart and after 30 Myr form an elongated structure almost 7 kpc long.

other for a long time–they get dispersed on time-scales on the order of few times 10^7 yr. We demonstrate this in Figure 3.12, which shows a series of snapshots around $z \approx 1.8$. In the first snapshot, marked with '$t = 0$', one clump starts to disintegrate. We indicate the positions of star particles which were formed in that clump by black asterisks. In the consecutive images, we show the positions of the same particles at later times. After ~ 30 Myr, stars which were formed in the same clump are distributed along an arc (or a spiral arm-like feature) ~ 7 kpc long.

Contribution to bulge growth

One of open questions related to giant clumps in high-redshift galaxies is what is their role in the build-up of the galactic bulge. The `m13` galaxy has a substantial spheroidal component at low redshifts, and its morphology at $z = 0$ is that of an elliptical galaxy, with stellar mass of $\sim 10^{11}$ $M_\odot$. As indicated before, the total mass of stars formed in clumps between $z \approx 2$ and $z \approx 1$ is $\sim 10^{10}$ $M_\odot$. Even if all that

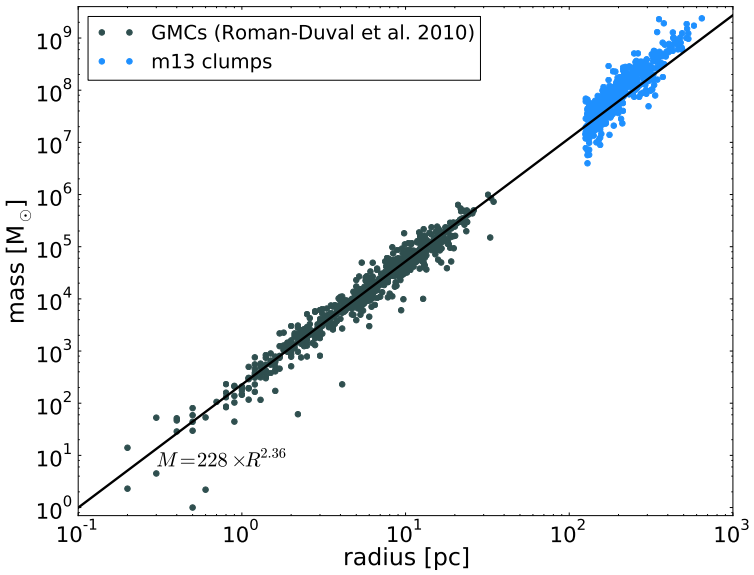

Figure 3.13: Comparison of giant clumps in our simulated high-redshift galaxy and GMCs in the Milky Way. The black line represents the mass-radius relation found for local GMCs by Roman-Duval et al. (2010). Giant clumps identified in the original series of snapshots of `m13` are shown in blue. The simulated giant clumps lie on an extension of the Milky Way GMC relation. The sharp cut-off in radius is a result of the minimum area threshold imposed by our clump-finding procedure.

mass ended up in the center of the galaxy, it would not be enough to account for the entire present-day bulge mass. However, even that is unlikely, as our results do not support the picture in which all clumps migrate to the center of the galaxy. We find that clumps do not experience very dramatic changes in their orbital angular momentum (up to $\sim 50\%$), and when they do change their radial position, they are equally likely to move inwards as they are to move outwards.

The results of our simulation therefore do not support the scenario in which giant star-forming clumps form bound stellar clusters in the outer regions of the disk which uniformly sink to the center of the galaxy, thus building up the bulge. Clumpy morphology of high-redshift galaxies may still play a role in bulge growth by facilitating transport of material across the disk via tidal forces and gravitational torquing. However, that analysis is beyond the scope of this study.

3.5.3 Clumps in comparison with GMCs

Finally, we compare giant clumps found in our simulated galaxy with the GMCs observed in the Milky Way. Local GMCs follow a tight relation between their

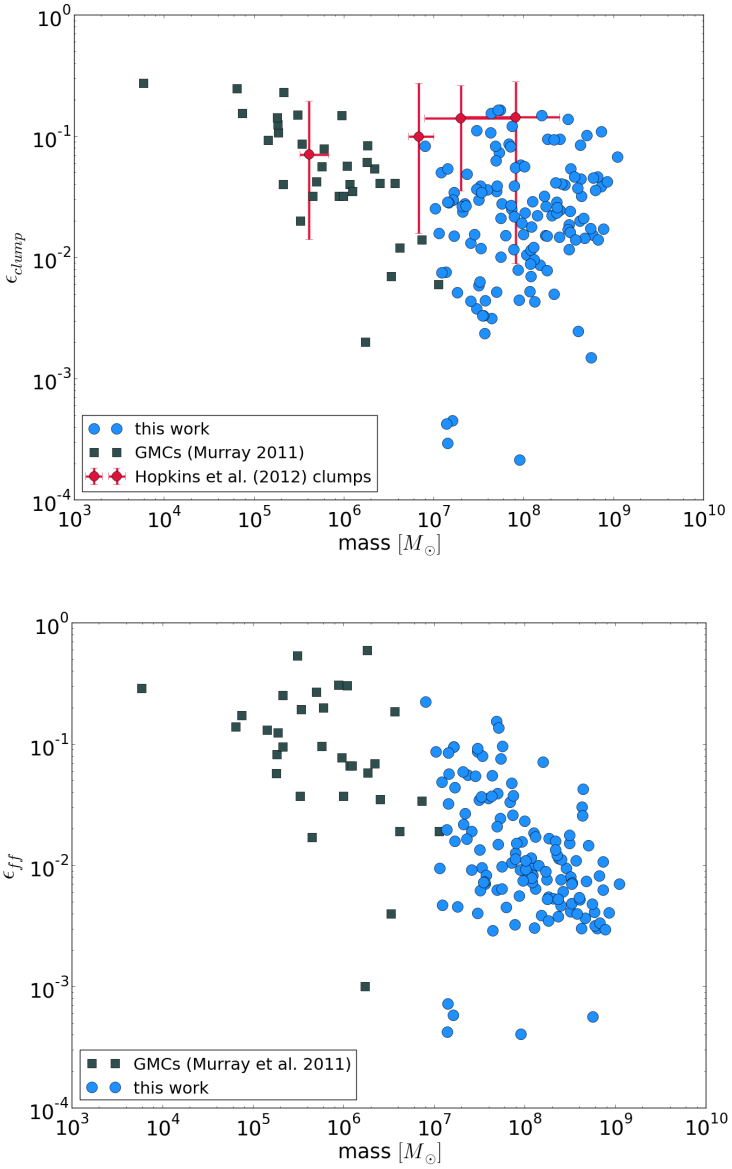

Figure 3.14: Clump star formation efficiency (ϵ_{clump}, upper panel) and star-formation efficiency per free-fall time (ϵ_{ff}, lower panel) as functions of mass, compared to the local GMCs (Murray 2011). In the upper panel, we also show the median value and 1σ range for clumps found in four high-resolution simulations of isolated galaxies by Hopkins et al. (2012).

mass and radius (Larson 1981), empirically derived to be M[$M_\odot$] = 228 $R^{2.36}$[pc] (Roman-Duval et al. 2010). Figure 3.13 shows that our simulated clumps follow that relation reasonably well, even though their masses and radii are orders of magnitude above those of the GMCs which were used to derive the relation. The sharp cut-off in radius of our clumps is a consequence of our clump identification criteria–we impose a minimum area of clumps of 5×10^4 pc^2, which corresponds to a minimum effective radius of $\sim$ 125 pc.

Next, we compare the star-formation efficiency (ϵ_{clump}) and star-formation efficiency per free-fall time (ϵ_{ff}) of giant clumps to those of local GMCs (Murray 2011). To calculate the star-formation efficiency of giant clumps, we use clumps found in six re-runs of our simulation in which we can follow the evolution of individual clumps from snapshot to snapshot. We compute ϵ_{clump} by calculating the total mass of stars formed in that clump over the course of its lifetime and dividing it by its maximum mass (i.e. total mass of gas and stars in the clump at the time when the clump is at its maximum mass):

$$\epsilon_{clump} = \frac{M_{*,formed}}{M_{clump}} \, . \tag{3.6}$$

To calculate the star formation efficiency per free-fall time ϵ_{ff}, which is more easily obtained in observations than ϵ_{clump}, we multiply ϵ_{clump} by the ratio of clump's free-fall time (τ_{ff}) and its lifetime. The free-fall time of a clump is given by

$$\tau_{ff} = \sqrt{\frac{3\pi}{32G\overline{\rho}}} \, , \tag{3.7}$$

where $\overline{\rho}$ is the mean gas density of the clump calculated at the time when the clump is at its maximum mass.

The results are shown in Figure 3.14. Clumps span a broad range in star-formation efficiency, from less than 10^{-3}, to a few tens of percent, indicating that giant clumps get destroyed well before they turn all their gas into stars. These values are very similar to star formation efficiency observed in the local GMCs (Murray 2011). They are also roughly consistent with the results from four simulations of isolated disk galaxies (Hopkins et al. 2012) which implement similar feedback processes as this work, but at higher resolution. Note that we use a different clump-finding algorithm and a different definition of clump lifetime than Hopkins et al. (2012) did, which may cause a variation, or perhaps a systematic offset, between our results.

Theoretical models by Krumholz & Dekel (2010) and Dekel & Krumholz (2013) predict that giant clumps with $M \sim 10^9$ $M_\odot$ and characterized by $\epsilon_{ff} \sim 1\%$ should

be stable and long-lived, which disagrees with our results. Their analytic models differ from our simulations not only in terms of complexity and the ability to treat non-linear effects, but also in assumptions regarding the nature of clumps–they assume that clumps are gravitationally bound objects, whereas we find that clumps are better described as loose complexes of bound sub-clumps (which are easier to disrupt)–and in the way stellar feedback is implemented (our feedback model gives a higher net momentum flux from stars).

3.6 Conclusions

We use high-resolution cosmological hydrodynamical simulations from the FIRE project which implement explicit models of stellar feedback and ISM physics to investigate giant star-forming clumps in high-redshift galaxies. We present a detailed analysis of a massive ($M_* \sim 10^{10.8}$ M$_\odot$ at $z = 1$), disky, star-forming galaxy over the redshift range $2.2 \geqslant z \geqslant 1.0$. Our main findings can be summarized as follows.

- At any given time around $z \sim 2$, the galaxy hosts a few giant clumps with baryonic masses in the range $\sim 10^8 - 10^9$ M$_\odot$ and radii $\sim 200 - 600$ pc. The clumps are identified as overdensities in the gas component of the disk; however, most of them also contain significant mass of stars (gas fraction in clumps ranges from 20% to over 90%). Clump shape can vary dramatically from clump to clump, from nearly spherical to highly elongated clumps.

- Maps of stellar surface density in general do not show prominent overdensities at the locations of gas clumps. A few prominent 'stellar clumps' found in our simulation belong to smaller galaxies merging with `m13`.

- The number of giant clumps decreases with time over the analysed redshift range, and the clumps become less massive at lower redshift. This follows the Toomre scale ($\sim f_g^3 M_{gal}$), given the decreasing galaxy gas fraction over the same time.

- Individual clump lifetimes are relatively short, typically $\lesssim 20 - 30$ Myr. More massive clumps on average live longer than less massive ones.

- Very massive clumps often end their lives by being disrupted from within and broken apart into smaller pieces by stellar feedback and tidal forces. Less massive clumps gradually fade away below the surface density threshold used to define clumps or merge with larger clumps.

- The mass-weighted stellar age of clumps ranges from $\lesssim$ 100 Myr to about 1 Gyr. This is much longer than the lifetime of any given gas clump, suggesting that a significant fraction of stars in a clump were formed before the clump itself. They most likely belong to the underlying stellar population of the galaxy which happens to spatially coincide with the gas clump. Therefore, stellar ages of clumps are not necessarily indicative of clumps' own lifetimes. At a given redshift, there is a tendency of clumps with older stellar populations to be located closer to the galactic center.

- Most clumps do not seem to be gravitationally bound, based on their often elongated and complex shape and the value of the virial parameter. Smaller regions within clumps are most likely bound and those are the regions where stars are born. The clumps are in fact 'complexes' with substantial substructure. This facilitates their destruction by stellar feedback. Stars formed in the same clump do not stay together for a long time; they drift apart on a time-scale of $\sim 10^7$ yr.

- During the gas-rich (or 'clumpy') phase in the evolution of the galaxy, clumps contain $\sim 10 - 60\%$ of the total star formation at any given time. They form $\sim 10^{10}$ $M_\odot$ of stars over the analysed redshift range, which is $\sim$20% of the total star formation during that period.

- Over the course of their short lifetime, clumps can either gain or lose orbital angular momentum. We find that both outcomes are roughly equally likely. There is no evidence of overall inward migration of clumps. Even the most dramatic angular momentum loss causes change only by a factor of ~ 2.

- The lack of evidence for systematic inward migration of clumps, and the total stellar mass produced in clumps, which is only a modest fraction ($\lesssim$ 20%) of the total $z = 0$ bulge mass, does not support the scenario in which giant star-forming clumps form bound stellar clusters which sink to the center of the galaxy, thus forming the bulge. Whether the mere presence of clumps helps the bulge growth by facilitating transport of material across the disk via tidal forces and gravitational torquing is an open question which is not addressed in this work.

- Giant clumps at high redshift look like scaled-up versions of GMCs in the local Universe. They lie on the same mass-radius relation and their star formation efficiencies are similar to those of GMCs.

APPENDIX

3.A Clump Properties in Additional FIRE Simulations of Massive Galaxies at Redshift 2

Simulation `m13` is the most massive galaxy in the original FIRE suite of simulations which were evolved all the way to $z = 0$. These simulations produce many galaxy properties which match observations across a broad redshift range (Hopkins et al. 2014). Simulation `m13` is the only galaxy in that sample with mass similar to the mass of observed clumpy galaxies at high redshift ($M_* \sim 10^{10} - 10^{11}$ $M_\odot$). To evolve such massive systems with high resolution down to $z = 0$ is computationally demanding, so `m13` was run at a lower resolution than the other original FIRE zoom-in simulations.

Because our analysis is based on one galaxy, there is a possibility that this system is a peculiar outlier from the rest of the population of similar galaxies. To test this, we compare the properties of `m13` at $z = 2.0$ with the properties of a couple of additional massive galaxies evolved down to $z \sim 2$, using the same code and the same implementation of stellar feedback, but at higher resolution. These galaxies (called *MFz2_A1* and *MFz2_A1*) are part of the *MassiveFIRE* sample (Feldmann et al. 2016), a suite of high-resolution simulations of massive galaxies at high-redshift ($z \gtrsim 1.7$). Their mass resolution is 16 times better (i.e. particle mass is 16 times smaller) than that of `m13`. Table 3.1 contains some of the basic properties of `m13` and the additional galaxies at $z = 2$.

These particular simulations were chosen because at $z \sim 2$ they do not show obvious signs of ongoing major mergers and appear to have disky morphologies. The gas surface densities of these two galaxies at $z = 2$ are shown in Figure 3.15. Upon further investigation, we noticed that in many $z \sim 2$ snapshots these galaxies do not have well-defined gas disks (unlike `m13` which has a substantial gas disk throughout the redshift range analysed in this work) and their gaseous material seems much more

Table 3.1: Physical parameters for the inner 10 kpc of `m13` and the additional MassiveFIRE galaxies, averaged over $\sim$ 150 Myr at $z \sim 2.0$.

name	M_* [$M_\odot$]	M_{gas} [$M_\odot$]	f_g	SFR [$M_\odot yr^{-1}$]
m13	3.8×10^{10}	1.6×10^{10}	0.29	44
MFz2_A1	2.3×10^{10}	0.4×10^{10}	0.13	5
MFz2_A3	1.1×10^{10}	0.6×10^{10}	0.32	4

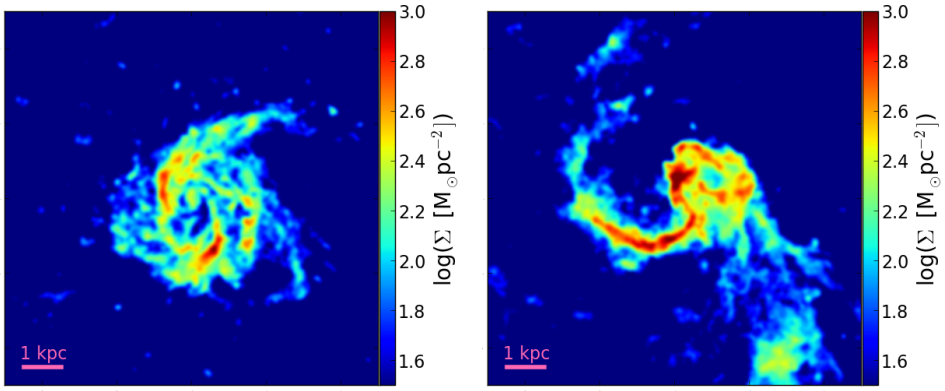

Figure 3.15: Gas surface density in the face-on projection at $z \sim 2$ for the two additional simulations, *MFz2_A1* in the left-hand panel and *MFz2_A3* in the right-hand panel.

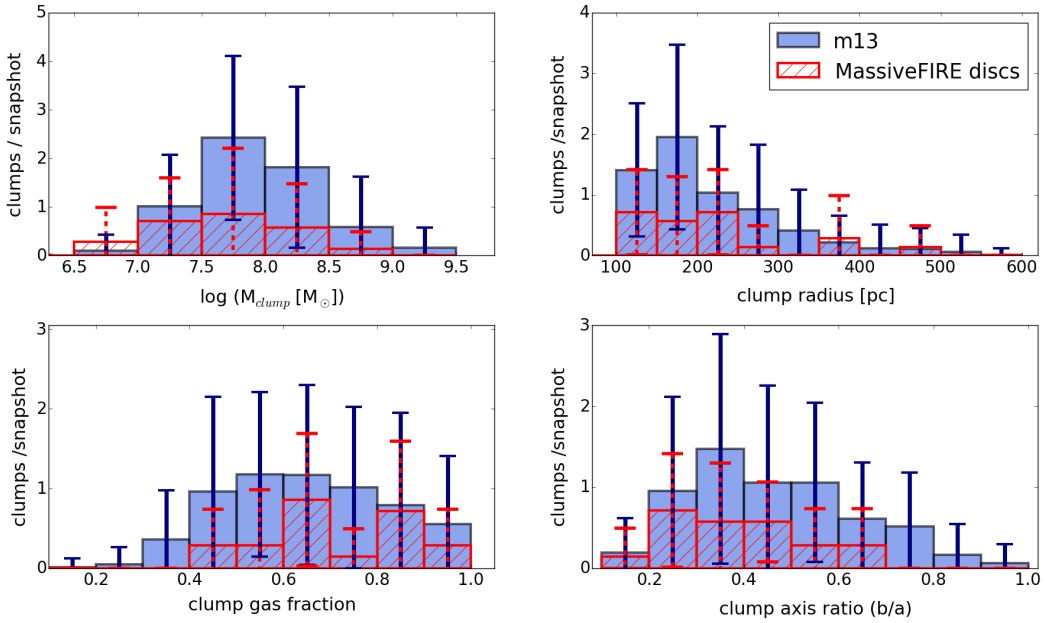

Figure 3.16: Distribution of clump properties, same as in Figure 3.3. Blue histograms show the average number of clumps in each bin, calculated using all clumps from the `m13` simulation. Error bars mark the standard deviation for each bin. The same is shown in red for clumps in $z \sim 2$ *MassiveFIRE* snapshots with disk-like gas distributions. The mean number of *MassiveFIRE* clumps in each bin falls within 1σ of the `m13` result.

erratic than that of m13. To make the comparison of clump properties more robust, we only use snapshots in which the additional simulated galaxies have disk-like gas distributions. This criterion however reduces our sample of 'good' snapshots to the total of seven (four from *MFz2_A3* and three from *MFz2_A1*).

We apply the clump-finding procedure described in Section 3.3 [8] to the snapshots of additional simulations and compare the clump properties to those of m13. The resulting distributions are shown in Figure 3.16. Blue histograms show the average number of clumps per bin for all m13 snapshots which contain gas clumps. The error bars show the standard deviation (1σ) for each bin. Results from the seven disk-like snapshots of *MassiveFIRE* simulations are shown in red. Even though we have a fairly small sample of suitable *MassiveFIRE* snapshots taken over a small redshift range, the average number of *MassiveFIRE* clumps per snapshot in each bin falls within 1σ of the m13 average, suggesting that clumps properties in the m13 and *MassiveFIRE* simulations are similar.

Although we do not include them in Figure 3.16, galaxies with non-disky gas distributions might be an interesting topic for future studies in the context of high-redshift clumpy galaxies. Applying the same analysis as for the snapshots with disk-like gas morphology, we found that some of these highly irregular gas distributions break up into gas clumps. These clumps might be a result of outflows and winds, unlike clumps in disks which most likely form via gravitational instability. Preliminary analysis of clumps in non-disky snapshots indicates that they might be more gas-dominated compared to clumps found in disks. A detailed investigation into the properties of clumps in non-disky galaxies, and how they compare to clumps observed in high-redshift galaxies, is left for future work.

3.B Sensitivity to the Clump Finder Parameters

In this appendix we analyze the sensitivity of clump properties on the gas surface density threshold used to define clumps in our clump-finding procedure. There is no single definition of a galaxy clump which is universally adopted, although attempts have been made to come up with one (see e.g. Guo et al. 2015). Different groups use different criteria to define clumps and to measure their properties. Here we show

[8] We had to slightly modify the way we compute the gas surface density threshold used in the clump-finding procedure to account for higher resolution (i.e. smaller particle mass) of *MassiveFIRE* simulations. In order to achieve the surface density threshold values in the *MassiveFIRE* snapshots similar to those we use for m13, in the high-resolution simulations we calculate the threshold using all cells with eight or more gas particles, instead of all cells with at least one gas particle.

that our choice of parameters used in the definition of clumps, although arbitrary to some extent, does not significantly affect the main result of our study.

The threshold value of the gas surface density which we use to define clumps is calculated for each snapshot as a 1σ above the mean surface density of non-empty grid cells in an *non-smoothed* 10-kpc map centered on `m13`. This definition of the threshold value was chosen after trying out many different options and evaluating each one by visual assessment. In other words, we chose the definition of clumps which seemed to work best at identifying regions of the disk which look like clumps in the maps of the gas surface density. This method is similar to what is done in observations, where clumps are usually identified by visual assessment.

In order to quantify how strongly our clumps are contrasted with respect to the rest of the disk, we compare the used threshold values to surface densities of smoothed maps on different scales. The used surface density threshold values correspond to regions which are overdense by factors of $\sim 10-15$ or $\sim 5\sigma$ above the mean surface density of pixels in the *smoothed* 10 kpc maps. However, if we consider a smaller region around the galaxy, both the mean and the standard deviation of the gas surface density increase and the nominal overdensity becomes lower. For example, our used threshold values correspond to a factor of $\sim 4-8$ or $\sim 3\sigma$ overdensities when the average values and sigmas are computed on a smoothed 5 kpc map around the center of the galaxy. This ambiguity in the density contrast of clumps is expected to be present not only in simulations, but also in observations of high-redshift galaxies, in which the definition of a clump may change depending on the sensitivity of observations.

To test whether a small change in the clump definition would have a significant impact on the derived properties of clumps, we re-analyze the original snapshots in the redshift range $2.0 \geq z > 1.8$ (20 snapshots in total) with different clump-finding criteria. We change the parameters of our clump finder in order to identify clumps which are 4, 5, and 6 standard deviations above the mean value, in the smoothed 10 kpc maps of the gas surface density. The average difference between the threshold values used in this comparison is about 18% [i.e. 5σ threshold values are higher (lower) than the 4σ (6σ) values by $\sim 18\%$].

Figure 3.17 compares the 5σ baryonic mass of each clump to the mass of the corresponding clump identified with the 4σ and 6σ definitions, shown in red and blue, respectively. For most clumps, especially the most massive ones, the total mass varies by less than a factor of two between the 5σ definition and the 4σ and

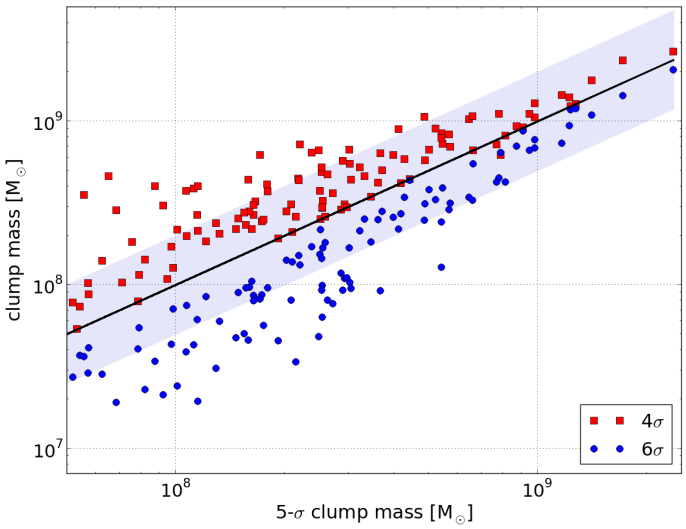

Figure 3.17: Comparison of clump mass found using different parameters in the clump-finding algorithm. Masses of clumps defined as 4σ and 6σ overdensities in the smoothed gas surface density maps are shown against the mass of the same clump found using the 5σ definition. If the masses were equal, they would fall on the black line. The colored band marks the region where the 4σ and 6σ masses are within a factor of two from the 5σ mass of the same clump. Most clumps populate this region, indicating that our estimates of clump mass are fairly robust to changing the clump definition.

6σ definitions.

Figure 3.18 shows how distributions of different clump properties change with changing the clump definition. Properties like the minor-to-major axis ratio and the gas fraction change very little. The distribution of clump radii seems to be more affected by changing the clump definition; however, the median values remain within a few tens of percent. In Figure 3.19, we compare the virial parameter of clumps found using these three definitions, and we find that they mostly overlap. Therefore, changes in the clump definition do not affect the (un-)bound state of clumps in a significant way.

In conclusion, although specific properties of any individual clump may change with changing the clump definition, these tests suggest that our main findings are not sensitive to small changes in the clump definition.

References

El-Badry, K., Wetzel, A., Geha, M., et al., 2016, ApJ, 820, 131, 131

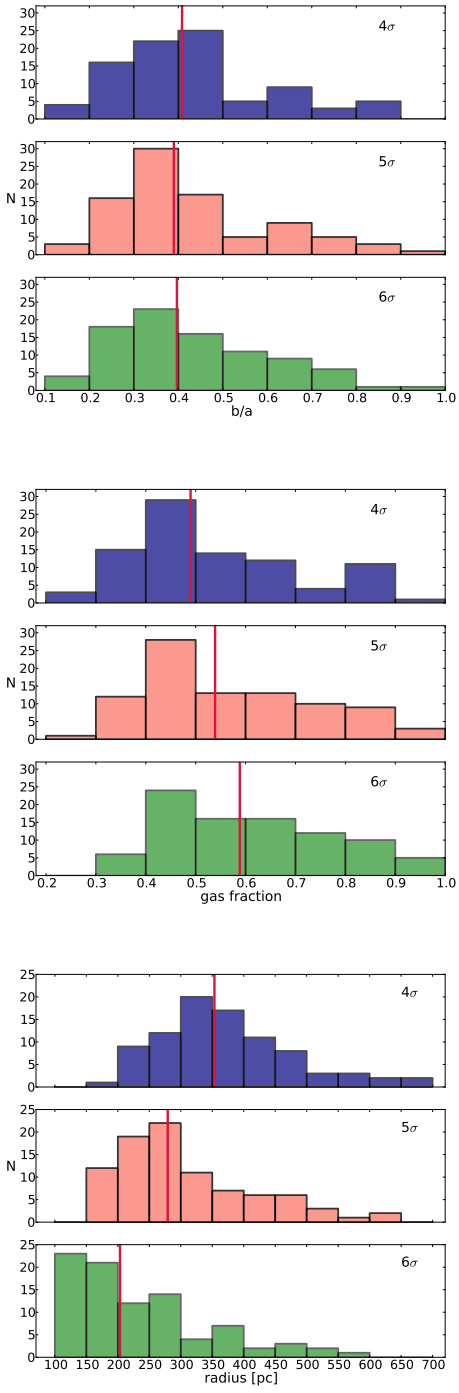

Figure 3.18: Distributions of clump properties (minor-to-major axis ratio, gas fraction, and clump radius) resulting from using different parameters (4σ, 5σ, and 6σ thresholds in the gas surface density) in the clump definition. Red lines show the median values for each case.

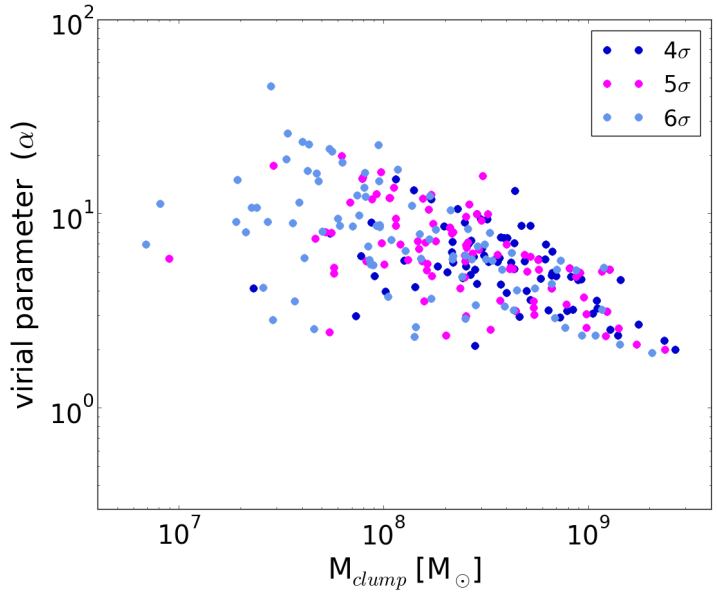

Figure 3.19: Virial parameter of a clump as a function of its mass, found using three different parameters in the clump-finding procedure (gas surface density thresholds set to 4σ, 5σ and 6σ above the mean value). There is no systematic difference in the clump virial parameter caused by using different definitions.

Adamo, A., Östlin, G., Bastian, N., et al., 2013, ApJ, 766, 105, 105

Agertz, O., Moore, B., Stadel, J., et al., 2007, MNRAS, 380, 963

Behrendt, M., Burkert, A., & Schartmann, M., 2016, ApJL, 819, L2, L2

Bertoldi, F., & McKee, C. F., 1992, ApJ, 395, 140

Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Bournaud, F., Dekel, A., Teyssier, R., et al., 2011, ApJL, 741, L33, L33

Bournaud, F., Perret, V., Renaud, F., et al., 2014, ApJ, 780, 57, 57

Brooks, A. M., Governato, F., Quinn, T., Brook, C. B., & Wadsley, J., 2009, ApJ, 694, 396

Ceverino, D., Dekel, A., & Bournaud, F., 2010, MNRAS, 404, 2151

Ceverino, D., Dekel, A., Mandelker, N., et al., 2012, MNRAS, 420, 3490

Cowie, L. L., Hu, E. M., & Songaila, A., 1995, AJ, 110, 1576

Dekel, A., Birnboim, Y., Engel, G., et al., 2009a, Nature, 457, 451

Dekel, A., Sari, R., & Ceverino, D., 2009b, ApJ, 703, 785

Dekel, A., & Krumholz, M. R., 2013, MNRAS, 432, 455

Elmegreen, B. G., 2011, ApJ, 737, 10, 10

Elmegreen, D. M., Elmegreen, B. G., Ravindranath, S., & Coe, D. A., 2007, ApJ, 658, 763

Elmegreen, B. G., Elmegreen, D. M., Fernandez, M. X., & Lemonias, J. J., 2009, ApJ, 692, 12

Faucher-Giguère, C.-A., Kereš, D., & Ma, C.-P., 2011, MNRAS, 417, 2982

Faucher-Giguère, C.-A., Quataert, E., & Hopkins, P. F., 2013, MNRAS, 433, 1970

Faucher-Giguère, C.-A., Hopkins, P. F., Kereš, D., et al., 2015, MNRAS, 449, 987

Faucher-Giguère, C.-A., Feldmann, R., Quataert, E., et al., 2016, MNRAS, 461, L32

Feldmann, R., Hopkins, P. F., Quataert, E., Faucher-Giguère, C.-A., & Kereš, D., 2016, MNRAS, 458, L14

Förster Schreiber, N. M., Shapley, A. E., Genzel, R., et al., 2011, ApJ, 739, 45, 45

Genel, S., Naab, T., Genzel, R., et al., 2012, ApJ, 745, 11, 11

Genzel, R., Burkert, A., Bouché, N., et al., 2008, ApJ, 687, 59

Genzel, R., Newman, S., Jones, T., et al., 2011, ApJ, 733, 101, 101

Guo, Y., Giavalisco, M., Ferguson, H. C., Cassata, P., & Koekemoer, A. M., 2012, ApJ, 757, 120, 120

Guo, Y., Ferguson, H. C., Bell, E. F., et al., 2015, ApJ, 800, 39, 39

Hahn, O., & Abel, T., 2011, MNRAS, 415, 2101

Hopkins, P. F., & Hernquist, L., 2010, MNRAS, 402, 985

Hopkins, P. F., Croton, D., Bundy, K., et al., 2010, ApJ, 724, 915

Hopkins, P. F., Quataert, E., & Murray, N., 2011, MNRAS, 417, 950

Hopkins, P. F., Kereš, D., Murray, N., Quataert, E., & Hernquist, L., 2012, MNRAS, 427, 968

Hopkins, P. F., Cox, T. J., Hernquist, L., et al., 2013, MNRAS, 430, 1901

Hopkins, P. F., Kereš, D., Oñorbe, J., et al., 2014, MNRAS, 445, 581

Immeli, A., Samland, M., Westera, P., & Gerhard, O., 2004, ApJ, 611, 20

Inoue, S., Dekel, A., Mandelker, N., et al., 2016, MNRAS, 456, 2052

Jog, C. J., 1996, MNRAS, 278, 209

Jones, T. A., Swinbank, A. M., Ellis, R. S., Richard, J., & Stark, D. P., 2010, MNRAS, 404, 1247

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R., 2005, MNRAS, 363, 2

Kereš, D., & Hernquist, L., 2009, ApJL, 700, L1

Kim, J.-h., Abel, T., Agertz, O., et al., 2014, ApJS, 210, 14, 14

Kroupa, P., 2002, Science, 295, 82

Krumholz, M. R., & Dekel, A., 2010, MNRAS, 406, 112

Larson, R. B., 1981, MNRAS, 194, 809

Leitherer, C., Schaerer, D., Goldader, J. D., et al., 1999, ApJS, 123, 3

Livermore, R. C., Jones, T., Richard, J., et al., 2012, MNRAS, 427, 688

Livermore, R. C., Jones, T. A., Richard, J., et al., 2015, MNRAS, 450, 1812

Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al., 2016, MNRAS, 456, 2140

Mandelker, N., Dekel, A., Ceverino, D., et al., 2014, MNRAS, 443, 3675

Mandelker, N., Dekel, A., Ceverino, D., et al., 2016, MNRAS

Moody, C. E., Guo, Y., Mandelker, N., et al., 2014, MNRAS, 444, 1389

Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al., 2015, MNRAS, 454, 2691

Murray, N., 2011, ApJ, 729, 133, 133

Murray, N., Quataert, E., & Thompson, T. A., 2010, ApJ, 709, 191

Noguchi, M., 1999, ApJ, 514, 77

Ostriker, E. C., & Shetty, R., 2011, ApJ, 731, 41, 41

Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R., 2010, ApJ, 723, 492

Somerville, R. S., Primack, J. R., & Faber, S. M., 2001, MNRAS, 320, 504

Springel, V., 2005, MNRAS, 364, 1105

Stewart, K. R., Bullock, J. S., Barton, E. J., & Wechsler, R. H., 2009, ApJ, 702, 1005

Swinbank, A. M., Smail, I., Longmore, S., et al., 2010, Nature, 464, 733

Tacconi, L. J., Genzel, R., Smail, I., et al., 2008, ApJ, 680, 246

Tacconi, L. J., Genzel, R., Neri, R., et al., 2010, Nature, 463, 781

Tacconi, L. J., Neri, R., Genzel, R., et al., 2013, ApJ, 768, 74, 74

Tamburello, V., Mayer, L., Shen, S., & Wadsley, J., 2015, MNRAS, 453, 2490

Toomre, A., 1964, ApJ, 139, 1217

van den Bergh, S., Abraham, R. G., Ellis, R. S., et al., 1996, AJ, 112, 359

Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al., 2012, ApJ, 753, 114, 114

Part II

Raman Scattering in Exoplanetary Atmospheres

Chapter 4

INTRODUCTION TO PART II

More than two decades have passed since the discovery of the first extrasolar planets (three pulsar planets PSR B1257+12 A, B, C; Wolszczan & Frail 1992; Wolszczan 1994) and the first planet found orbiting around a main-sequence star other than the Sun (51 Peg b; Mayor & Queloz 1995). Since then, almost 3,500 exoplanets have been discovered[1]. We now know that exoplanets are ubiquitous and that exoplanetary systems are remarkably diverse in terms of planets' masses, sizes, and orbital separations from their host stars (Winn & Fabrycky 2015).

In the decades prior to the discovery of extrasolar planets, theories of planetary formation and evolution have been designed and fine-tuned to explain the properties of the Solar System. One of the most interesting findings of the *Kelper* mission is the prevalence of planets with sizes between that of Earth and Neptune (the so-called super-Earths or mini-Neptunes) and with orbital periods of weeks to months (Batalha 2014), as shown in Figure 4.1. This seems to be a very common outcome of planet formation around solar-type stars, yet we had no knowledge of such objects prior to the discovery of exoplanets because there is no analog of such a planet in our Solar System. Now that we know that the Solar System's configuration is in no way preferred and not even very common, theoretical models of planetary formation and evolution need to be revisited and expanded.

Learning about the properties of exoplanets differs in many ways from research on the Solar System planets. Our knowledge about the properties of most planets in the Solar System is heavily based on measurements conducted by spacecraft sent to fly-by, orbit or land on these objects. For studying exoplanets, this is simply not possible. In order to gain insight into the properties of an exoplanet beyond its mass, radius, and some basic orbital parameters, we must rely on the information available through remote sensing of the planet's atmosphere. Using spectroscopic diagnostics we can, in principle, obtain information about atmospheric composition, temperature and pressure profiles, the presence of aerosols, the structure of winds and circulation patterns, and many other physical and chemical properties. Comprehensive characterization of exoplanet atmospheres is hence crucial for improving

[1] *NASA Exoplanet Archive*, April 2017.

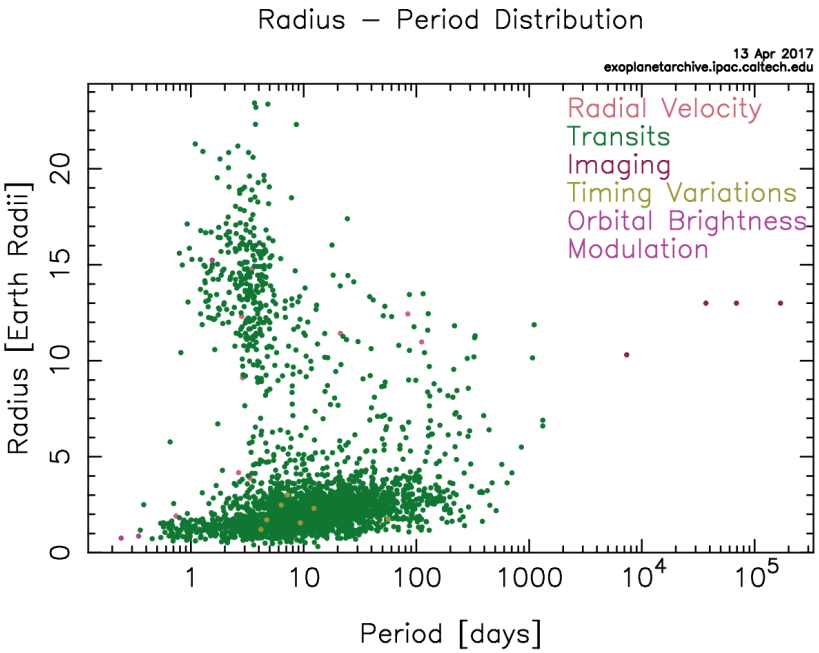

Figure 4.1: The radii and orbital periods of confirmed extrasolar planets, color-coded according to the method of their discovery. This figure was produced and made available by the *NASA Exoplanet Archive*.

our understanding of the exoplanet demographics and the physical processes that govern planetary formation and evolution.

4.1 Studies of Exoplanet Atmospheres

4.1.1 Observational techniques

The first detection of an exoplanet atmosphere was performed using transit spectroscopy of hot Jupiter HD 209458 b (Charbonneau et al. 2002) and most atmospheric detections to date have been accomplished using this observational technique. Planetary transits occur when planets whose orbital plane is aligned with our line of sight pass in front of the star, as seen from our point of view. If a planet has an atmosphere whose opacity varies with wavelength, then the effective radius of the planet (i.e., the radius at which the planet's atmosphere becomes optically thick) also changes with wavelength. By recording the effective size of the planet as a function of wavelength, the so-called transmission spectrum of a planet is obtained. By comparing the observed spectra with theoretical models, we can obtain information about the physical conditions in the atmosphere and the presence of different atomic and molecular species (Seager & Sasselov 2000; Hubbard et al. 2001; Brown 2001).

Half an orbit later, the secondary eclipse takes place – the planet passes behind the

star and the light reflected or emitted by the planet is blocked by the star. This results in a measurable reduction of the total light from the star-planet system. By comparing the observed signal during the eclipse (coming from the star alone) with that recorded just before or after the eclipse (containing light from both the star and the planet), the planetary signal can be inferred (Charbonneau et al. 2005; Deming et al. 2005).

The transiting observational methods are only applicable to those planets whose orbits are favorably oriented with respect to our line of sight. A large fraction of exoplanets is unobservable in this way. Fortunately, there are other techniques that can be used for detecting and characterizing exoplanetary atmospheres which do not require such fortuitous alignment of the planetary orbit.

High-resolution spectroscopy (with resolving power $R \gtrsim 30,000$) of large short-period planets is a technique in which the planet-star system is treated as a spectroscopic binary. The large radial velocity variations (on the order of km s^{-1}) in the planetary spectrum is used to distinguish the signal form the planet from the stellar spectrum (with typical radial velocities on the order of m s^{-1}) and from the contamination by the (stationary) telluric lines originating in the Earth's atmosphere. By cross-correlating the high-resolution planetary spectrum and molecular templates, the presence of different molecular species in the atmosphere can be inferred (Snellen et al. 2010; Crossfield et al. 2011; Brogi et al. 2012; Rodler et al. 2012; Birkby et al. 2013; de Kok et al. 2013; Lockwood et al. 2014). In order for this method to be successful at identifying molecules, highly reliable spectral line lists must exist, which is not always the case, especially at high temperatures relevant for the atmospheres of close-in planets. However, there has been a lot of effort in recent years on improving the spectroscopic data for important molecules in exoplanet atmospheres (Tennyson & Yurchenko 2012). In addition to atmospheric composition, the high-resolution spectroscopy method can be used to infer the planet's rotational velocity and the speed of the day-to-night-side winds (Snellen et al. 2010; Snellen et al. 2014; Brogi et al. 2016), as well as the orbital inclination of the planet (e.g. Birkby et al. 2017).

Observational methods discussed so far are based on disentangling the signal form the planet form observations of combined light of the planet and its host star. Direct imaging, on the other hand, is achieved by spatially resolving the planet from its host star and recording the signal from the planet alone. About two dozen substellar companions (i.e. planets and/or brown dwarfs) have been imaged since the first

direct detection more than an decade ago (e.g. Chauvin et al. 2005; Lagrange et al. 2010; Marois et al. 2008; Marois et al. 2010). Direct imaging is currently restricted to observations of young, massive gas giant planets located at large distances (tens of astronomical units) from their host stars. Young planets are hot and intrinsically bright because of the residual internal heat produced during their formation. Older planets have had time to cool down and their expected signal is hence dominated by the reflected starlight, instead of their own thermal emission. The flux coming from planets located close to their parent stars is usually buried in the noise of the diffracted stellar light. A great technical challenge in direct imaging is overcoming the large contrast between the planetary and stellar flux of many orders of magnitude (for example, the contrast between the Earth and the Sun in visible light is ten orders of magnitude). Diffraction-control systems such as coronagraphs, are often used to reduce that problem. In order to achieve good angular resolution, observations are usually made on large ground-based telescopes. The use of adaptive optics helps with reducing the effects of atmospheric scattering.

4.1.2 Prevalence of Exoplanet Atmospheres

All observational techniques used for atmospheric characterization that were discussed in the previous section work best for large and bright planets. Gas giants orbiting very close to their host stars (hot Jupiters) are particularly favorable targets for atmospheric characterization. For that reason, the majority of exoplanet atmospheres studied to date belong to gas giant exoplanets. Our knowledge about the atmospheres of smaller worlds is very limited at the moment. All large planets have thick atmospheres made of mostly hydrogen and helium; smaller, rocky planets may or may not have atmospheres. The transition between planets that must have atmospheres and those that may not, happens around planetary radii of $\sim 1.5 - 1.75\ R_\oplus$ (Lopez & Fortney 2014; Rogers 2015). The nature of atmospheres on the intermediate-sized worlds, the so-called super-Earths or mini-Neptunes (with masses up to $\sim 10\ M_\oplus$), is currently an open question and determining their properties is one of the most interesting and the most challenging problems in exoplanet science today. A wide range of planetary compositions can reproduce the average density of planets the size of Neptune or smaller (Adams et al. 2008). For example, a planet composed predominantly of water ice may have the same average density as a planet with a large rocky/iron core and a hydrogen/helium atmosphere. Super-Earths might have substantial hydrogen/helium atmospheres, or they may be more similar to terrestrial planets, with atmospheres made up of heavy molecules like

water or carbon dioxide (Miller-Ricci et al. 2009; Benneke & Seager 2013). To make things more uncertain, complex physical processes like atmospheric escape and outgassing, may completely alter the composition and the extent of a planetary atmosphere during the planet's lifetime (Howe & Burrows 2015).

4.1.3 Compositions

Early models of transmission spectra (Seager & Sasselov 2000) predicted strong absorption signals at the wavelengths of atomic resonance lines. The first robust detection of an exoplanet atmosphere was made through spectrophotometric observations of HD 209485 at wavelengths corresponding to the sodium resonance doublet at 589.3 nm (Charbonneau et al. 2002). Since then, sodium and potassium lines have been observed in exoplanets using space and ground-based observatories (e.g. Redfield et al. 2008; Sing et al. 2008; Sing et al. 2011; Huitson et al. 2012). Several atomic lines at shorter (UV) wavelengths, such as the hydrogen Lyα line and the prominent UV lines of O I, CII, and Si III, have been observed in a few short-period exoplanets to date, indicating the presence of extended atmospheres (i.e. exospheres) and providing evidence for ongoing atmospheric escape in these highly irradiated worlds (Vidal-Madjar et al. 2003; Vidal-Madjar et al. 2004; Linsky et al. 2010).

Due to their moderate temperatures, planetary atmospheres are mostly composed of molecules. The exact composition and the relative fractions of different molecular species depend on many complex physical and chemical processes taking place during the planet formation and its subsequent evolution. We hope to improve our knowledge and understanding of these processes by measuring the abundances of different molecules for different types of exoplanets.

The primary constituent of most exoplanetary atmospheres studied to date is molecular hydrogen, H_2. However, its presence is revealed to us only indirectly because H_2 is a spectrally inactive molecule which does not have prominent spectral features in the optical and near-infrared part of the spectrum. Some of the planets that have both the radius and mass measurements have average densities that can only be explained with a significant fraction of H_2/He envelopes (Fortney et al. 2007). The presence of H_2 can also be inferred form broad spectral features like the Rayleigh scattering slope in optical and UV wavelengths (Lecavelier Des Etangs et al. 2008). However, the Rayleigh scattering signal can be caused by other atmospheric constituents (other molecules or aerosols), so it cannot be used to unambiguously constrain the

presence of H_2. Furthermore, there have been claims that even non-atmospheric effects, like starspots, can produce false Rayleigh-like signals (Oshagh et al. 2014). The lack of a permanent dipole moment in H_2 can be partially overcome in conditions of high pressure where the effects of collisions between molecules become increasingly more important. Collisions can induce a series of broad absorption features in the infrared bands (Borysow et al. 1997). The evidence for this collision-induced-absorption have been seen in spectra of Jupiter, Saturn, and some brown dwarfs. However, these spectral features have not yet been used to measure the relative abundance of H_2 (Crossfield 2015).

Water and carbon monoxide have been robustly detected in exoplanet spectra using several different observational techniques. Near-infrared absorption bands of water have been detected in transmission and secondary eclipse spectra of giant exoplanets (e.g. Deming et al. 2013; Huitson et al. 2013; Wakeford et al. 2013; Kreidberg et al. 2014; Kreidberg et al. 2015; Line et al. 2016), as well as a Neptune-sized planet HAT-P-11 b (Fraine et al. 2014). Using the high-resolution spectroscopy method, H_2O and CO have been detected in both transiting and non-transiting hot Jupiters (e.g. Snellen et al. 2010; Birkby et al. 2013; Brogi et al. 2012; Rodler et al. 2012; Lockwood et al. 2014; Birkby et al. 2017). By cross-correlating moderate-resolution spectra of directly-imaged massive planets HR 8799 b and c with molecular templates, Konopacky et al. (2013) and Barman et al. (2015) found evidence for the presence of water and carbon monoxide in their atmospheres, as well as indication for the presence of methane in HR 8799 b.

4.1.4 Clouds and Hazes

Clouds and hazes can form in all Solar System planets with significant atmospheres, and based on the observations over the last several years, they seem to be quite common in exoplanetary atmospheres as well. Clouds are atmospheric condensates that form when the pressure of a condensible gas exceeds the saturation pressure and the excess material condenses into liquid droplets or solid 'dust' particles. Hazes are photochemically produced particles suspended in the atmosphere.

Both clouds and hazes can significantly impact the observable properties of exoplanet atmospheres. They contribute to atmospheric opacity and thus diminish the intensity of spectral features of atomic and molecular absorbers in the atmosphere. As a result, transmission spectra appear more flat than in the case of a clear (non-cloudy) atmosphere. This introduces degeneracy between the abundance of the

absorbers and the distribution of clouds (e.g. Line et al. 2016). Furthermore, fairly flat transmission spectra are also produced by clear atmospheres of high mean molecular weight, due to their reduced scale height. High-quality spectra are hence required in order to distinguish cloudy atmospheres from those composed of heavy molecular species.

4.2 Goals of Part II of this Thesis

Even though a lot of progress has been made over the last several years on identification of individual atomic and molecular species in exoplanet atmospheres, the relative abundances of different species in any exoplanet atmosphere are relatively poorly constrained. One major challenge is the presence of clouds and hazes, as discussed in the previous section. Another source of degeneracy is the fact that some very common atmospheric constituents are spectrally inactive and do not easily reveal their presence (even in substantial amounts) via spectroscopic analysis.

The most abundant atmospheric gases in all large planets – H_2 and He – fall into this category, and so does N_2, the most abundant constituent of the Earth's atmosphere. Benneke & Seager (2013)[2] showed that two atmospheres of very different overall composition can produce almost identical transmission spectra (Figure 4.2). Even though 90% of one atmosphere (shown in red) is in N_2 and H_2, while the other (shown in blue) is completely devoid of these gases, the two atmospheres have the same mean molecular weight and the same abundance ratio of strong absorbers (water and CO_2) which completely dominate the spectrum, especially at near-infrared wavelengths. In order to break this degeneracy and reliably determine the atmospheric composition and relative abundances of all species, we must be able to detect and identify spectrally inactive gases.

This served as motivation for looking into the effects of Raman scattering on molecules in exoplanet atmospheres, which is the subject of the second part of this thesis. As discussed in more detail in Chapter 5, Raman scattering on molecules introduces specific spectral features in the reflected light spectra of planets. These spectral features carry the signature of the scattering molecule(s) and can be used to identify gases present in the atmosphere in large quantities, even spectrally inactive molecules like H_2 and N_2. Chapter 5 describes how the intensity and structure of these Raman features can be used to infer some important properties of the exoplanetary atmosphere, such as its composition, temperature, and whether the atmosphere

[2] dx.doi.org/10.1088/0004-637X/778/2/153

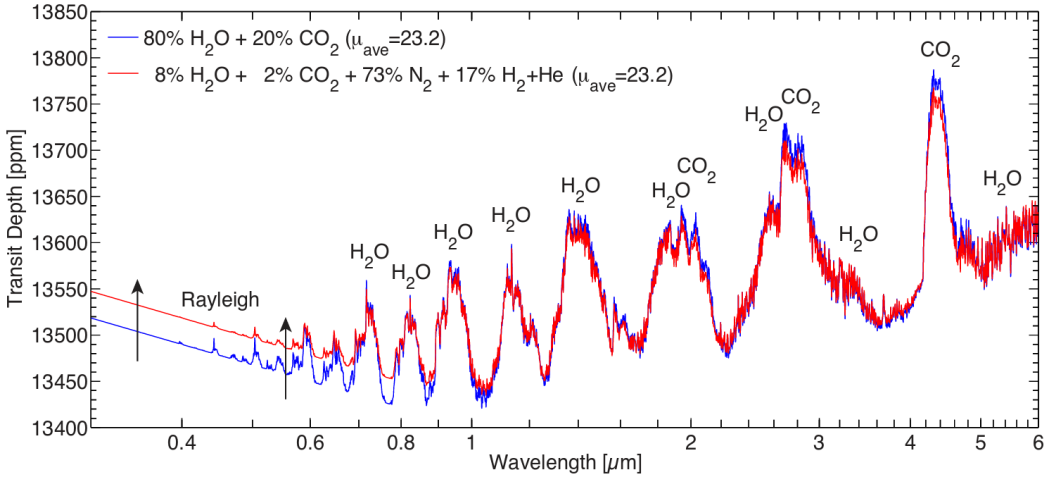

Figure 4.2: Model transmission spectra from Benneke & Seager (2013, Figure 8) for two atmospheres of the same mean molecular weight and the same relative abundance of strong absorbers, H_2O and CO_2, but vastly different overall compositions. These spectra demonstrate the challenges in distinguishing between atmospheres of different compositions due to the presence of spectrally inactive gases, such as H_2, He, and N_2. ©AAS. Reproduced with permission.

is clear or cloudy. I present a feasibility analysis for observing the Raman features in nearby exoplanets using the next-generation observational facilities. The shape and the intensity of Raman features also heavily depends on the details of the incident stellar spectrum. In Chapter 6, I expand upon my previous work by analyzing the diversity of Raman features in planets around various types of host stars.

Detecting the signatures of Raman scattering in exoplanet atmospheres is most likely beyond the reach of current instruments. However, the field of exoplanet science is rapidly advancing and it will greatly benefit from new observational facilities planned for the near and slightly more distant future. Two large upcoming space missions – the *James Webb Space Telescope (JWST)* and the *Wide-Field Infrared Survey Telescope (WFIRST)* – will not cover the spectral range in which Raman features are most prominent ($\lambda \lesssim 600$ nm). However, two concept missions are currently being studied for consideration by the 2020 Decadal Survey – the *Habitable Exoplanet Imaging Mission (HabEx)* and the *Large UV/Optical/Infrared Surveyor (LUVOIR)*. Both concepts cover a wide spectral range, reaching down to short visible/UV wavelengths. The *HabEx* mission design includes a coronagraph and/or a starshade in order to achieve extremely high contrast imaging, capable of detecting terrestrial planets orbiting nearby solar-type stars. Detecting the spectral signatures of Raman scattering should be within reach of these two telescopes, as

well as the next generation of large ($\sim 25 - 40$ m) ground-based observatories.

References

Adams, E. R., Seager, S., & Elkins-Tanton, L., 2008, ApJ, 673, 1160-1164, 1160

Barman, T. S., Konopacky, Q. M., Macintosh, B., & Marois, C., 2015, ApJ, 804, 61, 61

Batalha, N. M., 2014, Proceedings of the National Academy of Science, 111, 12647

Benneke, B., & Seager, S., 2013, ApJ, 778, 153, 153

Birkby, J. L., de Kok, R. J., Brogi, M., et al., 2013, MNRAS, 436, L35

Birkby, J. L., de Kok, R. J., Brogi, M., Schwarz, H., & Snellen, I. A. G., 2017, AJ, 153, 138, 138

Borysow, A., Jorgensen, U. G., & Zheng, C., 1997, A&A, 324, 185

Brogi, M., Snellen, I. A. G., de Kok, R. J., et al., 2012, Nature, 486, 502

Brogi, M., de Kok, R. J., Albrecht, S., et al., 2016, ApJ, 817, 106, 106

Brown, T. M., 2001, ApJ, 553, 1006

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L., 2002, ApJ, 568, 377

Charbonneau, D., Allen, L. E., Megeath, S. T., et al., 2005, ApJ, 626, 523

Chauvin, G., Lagrange, A.-M., Dumas, C., et al., 2005, A&A, 438, L25

Crossfield, I. J. M., 2015, PASP, 127, 941

Crossfield, I. J. M., Barman, T. S., & Hansen, B. M. S., 2011, ApJ, 736, 132, 132

de Kok, R. J., Brogi, M., Snellen, I. A. G., et al., 2013, A&A, 554, A82, A82

Deming, D., Seager, S., Richardson, L. J., & Harrington, J., 2005, Nature, 434, 740

Deming, D., Wilkins, A., McCullough, P., et al., 2013, ApJ, 774, 95, 95

Fortney, J. J., Marley, M. S., & Barnes, J. W., 2007, ApJ, 659, 1661

Fraine, J., Deming, D., Benneke, B., et al., 2014, Nature, 513, 526

Howe, A. R., & Burrows, A., 2015, ApJ, 808, 150, 150

Hubbard, W. B., Fortney, J. J., Lunine, J. I., et al., 2001, ApJ, 560, 413

Huitson, C. M., Sing, D. K., Vidal-Madjar, A., et al., 2012, MNRAS, 422, 2477

Huitson, C. M., Sing, D. K., Pont, F., et al., 2013, MNRAS, 434, 3252

Konopacky, Q. M., Barman, T. S., Macintosh, B. A., & Marois, C., 2013, Science, 339, 1398

Kreidberg, L., Bean, J. L., Désert, J.-M., et al., 2014, ApJL, 793, L27, L27

Kreidberg, L., Line, M. R., Bean, J. L., et al., 2015, ApJ, 814, 66, 66

Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al., 2010, Science, 329, 57

Lecavelier Des Etangs, A., Vidal-Madjar, A., Désert, J.-M., & Sing, D., 2008, A&A, 485, 865

Line, M. R., Stevenson, K. B., Bean, J., et al., 2016, AJ, 152, 203, 203

Linsky, J. L., Yang, H., France, K., et al., 2010, ApJ, 717, 1291

Lockwood, A. C., Johnson, J. A., Bender, C. F., et al., 2014, ApJL, 783, L29, L29

Lopez, E. D., & Fortney, J. J., 2014, ApJ, 792, 1, 1

Marois, C., Macintosh, B., Barman, T., et al., 2008, Science, 322, 1348

Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T., 2010, Nature, 468, 1080

Mayor, M., & Queloz, D., 1995, Nature, 378, 355

Miller-Ricci, E., Seager, S., & Sasselov, D. D., 2009, ApJ, 690, 1056

Oshagh, M., Santos, N. C., Ehrenreich, D., et al., 2014, A&A, 568, A99, A99

Redfield, S., Endl, M., Cochran, W. D., & Koesterke, L., 2008, ApJL, 673, L87, L87

Rodler, F., Lopez-Morales, M., & Ribas, I., 2012, ApJL, 753, L25, L25

Rogers, L. A., 2015, ApJ, 801, 41, 41

Seager, S., & Sasselov, D. D., 2000, ApJ, 537, 916

Sing, D. K., Vidal-Madjar, A., Désert, J.-M., Lecavelier des Etangs, A., & Ballester, G., 2008, ApJ, 686, 658-666, 658

Sing, D. K., Désert, J.-M., Fortney, J. J., et al., 2011, A&A, 527, A73, A73

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S., 2010, Nature, 465, 1049

Snellen, I. A. G., Brandl, B. R., de Kok, R. J., et al., 2014, Nature, 509, 63

Tennyson, J., & Yurchenko, S. N., 2012, MNRAS, 425, 21

Vidal-Madjar, A., Lecavelier des Etangs, A., Désert, J.-M., et al., 2003, Nature, 422, 143

Vidal-Madjar, A., Désert, J.-M., Lecavelier des Etangs, A., et al., 2004, ApJL, 604, L69

Wakeford, H. R., Sing, D. K., Deming, D., et al., 2013, MNRAS, 435, 3481

Winn, J. N., & Fabrycky, D. C., 2015, ARA&A, 53, 409

Wolszczan, A., 1994, Science, 264, 538

Wolszczan, A., & Frail, D. A., 1992, Nature, 355, 145

Chapter 5

RAMAN SCATTERING BY MOLECULAR HYDROGEN AND NITROGEN IN EXOPLANETARY ATMOSPHERES

ABSTRACT

An important source of opacity in exoplanet atmospheres at short visible and near-UV wavelengths is Rayleigh scattering of light on molecules. It is accompanied by a related, albeit weaker process – Raman scattering. We analyze the signatures of Raman scattering imprinted in the reflected light and the geometric albedo of exoplanets, which could provide information about atmospheric properties. Raman scattering affects the geometric albedo spectra of planets in following ways. Firstly, it causes filling-in of strong absorption lines in the incident radiation, thus producing sharp peaks in the albedo. Secondly, it shifts the wavelengths of spectral features in the reflected light causing the so-called Raman ghost lines. Raman scattering can also cause a broadband reduction of the albedo due to wavelength shifting of a stellar spectrum with red spectral index. Observing the Raman peaks in the albedo could be used to measure the column density of gas, thus providing constrains on the presence of clouds in the atmosphere. Observing the Raman ghost lines could be used to spectroscopically identify the main scatterer in the atmosphere, even molecules like H_2 or N_2 that do not have prominent spectral signatures in the optical wavelength range. If detected, ghost lines could also provide information about the temperature of the atmosphere. Here, we investigate the effects of Raman scattering in hydrogen- and nitrogen-dominated atmospheres. We analyze the feasibility of detecting the signatures of Raman scattering with the existing and future observational facilities, and of using these signatures as probes for exoplanetary atmospheres.

The material in this chapter has been published as Oklopčić, A., Hirata, C.M. & Heng, K. 2016, ApJ, 832, 30.

5.1 Introduction

Characterizing atmospheres of exoplanets usually involves detecting signatures of strong absorbers present in the atmosphere, such as alkali atoms sodium and potassium, and molecules like H_2O, CH_4, CO and CO_2 (e.g. Seager & Sasselov 2000; Brown 2001; Hubbard et al. 2001; Sudarsky et al. 2003). However, these atomic and molecular species can often be present in small amounts and the bulk of the atmosphere can be made of spectroscopically inactive species – molecular hydrogen and helium in the case of giant planets, or molecular nitrogen in the case of the Earth's atmosphere. Determining the composition of the spectroscopically inactive gas components is very important for atmosphere characterization. Constraining the abundance of bulk gases breaks the degeneracies associated with measuring the (relative) abundances of trace gases and other atmospheric properties, such as the atmospheric pressure and temperature which can affect the strength of absorption lines. However, detecting molecules like H_2 and N_2 directly (i.e. spectroscopically) is challenging because these are homonuclear diatomic molecules with no low-lying excited electronic states, and hence do not have prominent spectral features in the optical wavelength range.

At high densities, these molecules can form temporary collisional pairs that have non-zero electric dipole moments and therefore exhibit radiative dipole transitions. Via this so-called collision-induced absorption (CIA) mechanism, H_2-H_2 and H_2-He pairs become an additional source of opacity in the infrared part of the spectrum (Saumon et al. 2012) that can be used to constrain the abundance of molecular hydrogen and helium. Schwieterman et al. (2015) similarly proposed using an absorption feature of a collisional pair of N_2 molecules at 4.3 μm to detect and constrain the abundance of molecular nitrogen in the atmospheres of terrestrial exoplanets. These broad spectral features of H_2 and N_2 can be very useful probes of deep and dense atmospheres. However, the spectral range in which they reside (NIR and IR) contains numerous spectral lines/bands of other strong absorbers like methane, water, and carbon dioxide with which CIA features often overlap, which can lead to uncertainties in the measured abundances. Therefore, having a complementary method of detecting signatures of H_2 or N_2 in the visible part of the spectrum would be very useful for breaking some of these degeneracies.

An important source of opacity at short optical and UV wavelengths is Rayleigh scattering of radiation by abundant molecules like H_2 (Marley et al. 1999; Sudarsky et al. 2000; Hubbard et al. 2001). Rayleigh scattering has been observed in trans-

mission spectra of hot Jupiters (Lecavelier Des Etangs et al. 2008a; Lecavelier Des Etangs et al. 2008b); however this is not an unambiguous sign of H_2 because the observed spectral slope could also be produced by Rayleigh scattering by condensate particles, as discussed by Lecavelier Des Etangs et al. (2008a). Rayleigh scattering does not leave direct spectroscopic signatures such as distinct spectral lines that could be used to unambiguously identify the main scatterer.

A related, albeit weaker, process – Raman scattering – does imprint signatures of scattering molecules in the scattered light and hence in the albedo of planets. If these Raman features in the albedo were detected, they could be used to distinguish between molecular scattering and scattering by small particles which does not produce Raman features, thus providing information on the presence of clouds or hazes in the observed atmosphere. Additionally, Raman scattering features might be used to spectroscopically identify the dominant scattering molecule, such as H_2 or N_2, and even the relative population of different molecular levels, thus constraining the temperature of the medium.

The effects of Raman scattering on planetary albedos and its potential for probing planetary atmospheres has been known and studied for a few decades, mostly in the context of Solar System planets (for reviews of the early studies of Raman scattering see Price 1977; Cochran 1981). The purpose of this study is to demonstrate that detecting the signatures of Raman scattering in the reflected light from *exoplanets* could be a powerful tool in exoplanet atmosphere characterization and hence may become an interesting target for observations in the era of the next-generation observational facilities. Here, we investigate the effects of Raman scattering by molecular hydrogen and nitrogen on the geometric albedo spectra of exoplanets and analyze the feasibility of detecting these features in nearby exoplanets using the currently available and future observational facilities.

In Section 5.2 we give an overview of the basic physics related to Raman scattering and describe the effects of Raman scattering on the reflected light from planets and their albedo. In Section 5.3 we describe the radiative transfer calculations and model atmospheres that we use in this study. We present the results of radiative transfer calculations for two test cases – a monochromatic light source and a flat photon spectrum with an absorption line in Section 5.4. The goal of these test cases is to provide insight into the way in which Raman scattering affects the reflected light from planets. We then present the calculated albedo spectra for several models of clear and cloudy hydrogen- and nitrogen-dominated atmospheres irradiated by

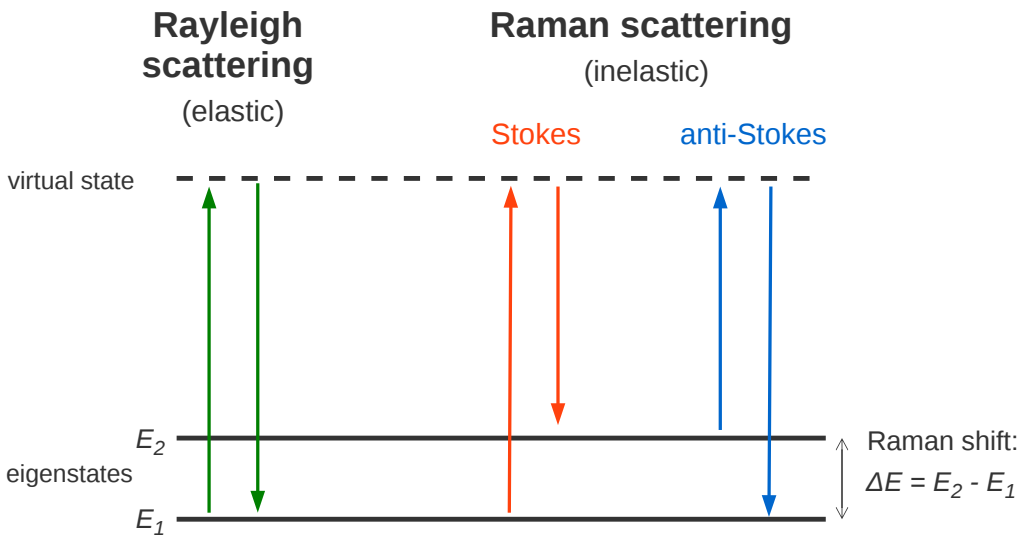

Figure 5.1: Schematic representation of Rayleigh and Raman scattering. The initial and final states are eigenstates of the molecule, whereas the intermediate state is virtual, making these processes sources of continuum opacity. The energy of the final state can be higher (Stokes Raman scattering), lower (anti-Stokes scattering), or equal to that of the initial state (Rayleigh scattering). The change in the photon energy (Raman shift) is determined by the structure of molecular energy levels, and is unique for every molecule.

the solar spectrum. We discuss our results and analyze the detectability of Raman features in the reflected light from exoplanets in Section 5.5. In Section 5.6 we present the summary and the conclusions of this work.

5.2 Raman Scattering

Scattering is an interaction between radiation and matter in which an incident photon is annihilated, and a new photon is created. If the scattered photon has the same energy as the incident one, the process is called Rayleigh scattering. If the energies of the incident and the scattered photon are different, we talk about Raman scattering. Raman and Rayleigh scattering are sources of continuum opacity because in both cases the intermediate state of the atom or the molecule is a virtual state, not an eigenstate of the system, as in the case of resonance scattering. For a detailed treatment of Raman scattering we point the reader to Long 2002.

Figure 5.1 shows schematically the mechanism of Raman scattering, as well as

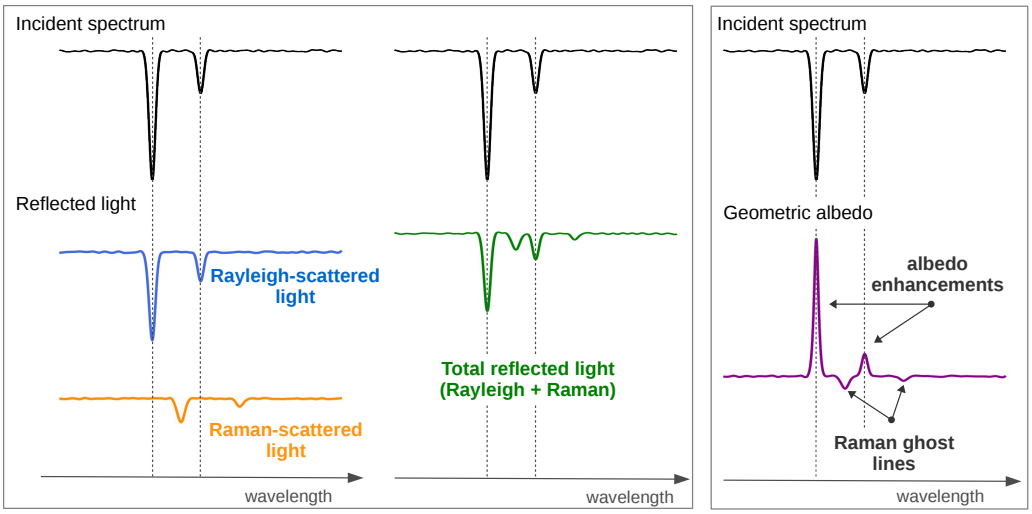

Figure 5.2: Reflected light due to Rayleigh scattering is proportional to the stellar flux at the same wavelength. The contribution from Raman scattering depends on the incident flux at the Raman-shifted wavelengths. If there are strong features such as absorption lines in the incident spectrum, they will be present in the reflected light at the same wavelength (due to Rayleigh scattering) and at shifted positions (due to Raman scattering). Consequently, there are two types of features in the albedo spectrum due to Raman scattering: i) albedo enhancements at the wavelengths of absorption lines due to filling-in of the lines by Raman-shifted light and ii) dips in the albedo at wavelengths of the so-called Raman ghosts of the absorption lines. A pure Rayleigh-scattering atmosphere (with no Raman scattering and no absorption) would result in a flat, featureless geometric albedo spectrum.

Rayleigh scattering, which may be thought of as a special case of Raman scattering. Depending on whether the scattered photon has lower or higher energy than the incident one, we distinguish between Stokes and anti-Stokes Raman processes. Due to Raman scattering the molecule changes its rotational and/or vibrational state. The allowed changes in the rotational quantum number of the molecule are $\Delta J \equiv J_f - J_i = -2, 0, 2$, corresponding to transitions with spectroscopic labels O, Q, and S, respectively. Vibrational quantum number of the molecule changes in the case of pure vibrational (Q-branch) or rotational-vibrational (O- and S-branch) Raman scattering. Raman transitions depend on the molecular polarizability, not on its electric dipole moment like the regular rotational and vibrational transitions. That is why even homonuclear diatomic molecules, like H_2 and N_2, that have zero dipole moments and hence do not exhibit rotational and ro-vibrational transitions, do exhibit Raman transitions, since their polarizability is non-zero.

The change in the energy of the scattered photon, i.e. the Raman shift, is determined

by the energy difference between the initial and the final state of the molecule. This property of Raman scattering can be used to spectroscopically identify the scattering medium by measuring the Raman shift that photons experience. Because the exact value of the Raman shift depends on the initial and the final state of the molecule and because the intensity of a shifted line is proportional to the population of the initial state (in the optically thin case), a detailed analysis of the Raman scattering features and their relative intensities can provide information about the relative population of different molecular levels, and hence be used to constrain the temperature of the scattering medium.

If the intensity of light reflected off a planet is directly proportional to the intensity of the incident stellar light at the same wavelength, as it is in the case of a deep (semi-infinite) Rayleigh-scattering atmosphere without any absorption, the resulting geometric albedo spectrum of the planet (the ratio of the reflected and incident light as a function of wavelength) is flat and featureless.[1] On the other hand, the intensity of Raman-scattered light at any wavelength depends on the intensity of the incident radiation at a different, Raman-shifted wavelength. Due to this additional component, the total reflected spectrum (comprising of Rayleigh and Raman contributions) is not simply proportional to the incident spectrum and the resulting albedo is not smooth – it contains spectral features characteristic of Raman scattering. In this work, we focus on two types of features that appear in the planetary albedo spectrum as a result of Raman scattering (see Figure 5.2 for reference):

1. *Albedo enhancements*: if the incident spectrum has some distinct spectral features, such as strong absorption lines, they will be present in the reflected light at the same wavelength due to Rayleigh scattering. However, the Raman-scattered component at that wavelength is related to the intensity of the incident radiation at the Raman-shifted wavelength, where there is (most likely) no line in the stellar spectrum and the incident intensity is higher. Therefore, the relative contribution of Raman-scattered light to the total reflected light is enhanced at wavelengths corresponding to prominent absorption lines, thus filling-in these lines (i.e. reducing their depth) in the reflected light.

   This additional component in the reflected light at the wavelengths of absorption lines in the stellar spectrum, compared to the surrounding continuum,

[1] For a Rayleigh-scattering atmosphere of finite depth the albedo spectrum is not necessarily flat, as shown in Sromovsky (2005a, Figure 2), but it is always a smoothly-varying function of optical depth, and consequently, of frequency. Therefore, sharp features like spectral lines do not appear in the albedo spectrum caused by Rayleigh scattering.

causes positive peaks in the albedo spectrum, which we call albedo enhancements. The effect is most pronounced for the strongest absorption lines in the incident spectrum.

2. *Raman ghosts*: the intensity of Raman-scattered light is reduced at wavelengths corresponding to Raman-shifted positions of strong absorption lines in the incident spectrum. Therefore, the total intensity of the reflected light at these wavelengths is reduced compared to the surrounding continuum, creating weak dips in the reflected spectrum, and consequently in the planet's albedo, called Raman ghosts.

The Raman effect occurs only on scattered photons; its signatures can only be observed in the reflected light, and not in transmission spectra of exoplanets.

The importance of Raman scattering in planetary atmospheres was first discovered in the 1960s. Brinkmann (1968) used it to explain the observed filling-in of solar Fraunhofer lines in the Earth's atmosphere, also known as the Ring effect (Grainger & Ring 1962). Studies of Raman scattering in the context of atmospheres of other planets began soon thereafter (Wallace 1972; Belton et al. 1971; Belton et al. 1973; Price 1977; Cochran & Trafton 1978; Cochran 1981; Cochran et al. 1981) and in the next few decades the effects of Raman scattering have been observed in all gas planets in the Solar System (see e.g. Karkoschka 1994; Yelle et al. 1987; Bétremieux & Yelle 1999; Courtin 1999).

To include the effects of Raman scattering in the atmospheric models, we follow the method presented in Bétremieux & Yelle (1999). They studied Raman scattering in the atmosphere of Jupiter, which they observed using the Hubble Space Telescope. By comparing the observations with models, they identified specific Raman features and used them to determine the rotational distribution of H_2 in the Jupiter's atmosphere and the H_2 ortho-para ratio. Their models were computed using the publicly available radiative transfer code DISORT (Stamnes et al. 1988), which they modified to include the effects of multiple Raman scattering by H_2. In this study, we also use DISORT, to which we have made similar modifications in order to include the treatment of Raman scattering.

For the purpose of studying the atmosphere of Neptune, Sromovsky (2005b) developed the first method for treating Raman scattering that includes polarization effects. This accurate method was then used to evaluate several approximate methods for including Raman scattering in radiative transfer calculations that were developed

with the goal of reducing computational costs associated with rigorous treatment of Raman scattering. The approximation introduced by Pollack et al. (1986) is often used to treat Raman scattering in modeling of reflected light and albedos of exoplanets (e.g. Marley et al. 1999; Sudarsky et al. 2000; Cahoy et al. 2010). Marley et al. (1999) use the Pollack approximation with correction terms calculated for a 6000-K blackbody. The albedo spectra produced by their models do not show distinct Raman features that are the focus of our study because a blackbody spectrum does not contain any lines.

5.3 Methods

5.3.1 Model Atmospheres

We investigate the signatures of Raman scattering imprinted on the geometric albedo of H_2- and N_2-dominated atmospheres. We focus our attention on the near-ultraviolet and blue end of the visible spectrum (the wavelength range $3000 \leq \lambda \leq 5000$ Å) because the Raman scattering cross section has a strong wavelength dependence ($\sigma \propto \lambda^{-4}$) and the effects that we investigate become weaker as the wavelength increases. Compared to longer wavelengths, this part of the spectrum is relatively clean of absorption lines from atoms and molecules present in the atmosphere. The most notable absorption lines are those of sodium and potassium at ~ 330 nm and ~ 404 nm, respectively. The abundances of Na and K atoms in monatomic form strongly depend on the presence and structure of clouds in the atmosphere. We simplify our model atmospheres by neglecting absorption and including only opacity due to (Rayleigh and Raman) scattering. This allows us to more easily identify and study the effects of Raman scattering alone, which is the main focus of our investigation.

For our main model atmosphere we assume that it is composed entirely of molecular hydrogen $^1H_2$ (90% by number) and helium ^4He (10%). Both hydrogen and helium contribute to Rayleigh scattering, but only hydrogen molecules Raman scatter in the wavelength range that we are interested in. The mean molecular weight of this atmosphere is set to $\mu = 2.3\ m_p$ and the surface gravity is $g = 10^3$ cm s^{-2}.

We use three different temperature-pressure (T-P) profiles for clear (cloud-free) H_2/He atmospheres:

1. Hot atmosphere: T-P profile of the cloud-free case from Heng et al. (2012). The temperature ranges from 1350 K to 1600 K.

2. Warm atmosphere: T-P profile of the cloud-free model for warm Neptunes from Cahoy et al. (2010) – the Neptune-like planet at 0.8 AU from the host star – with temperature in the range $\sim 300 - 1000$ K.

3. Cold atmosphere: T-P profile of Neptune from Sromovsky (2005b, Fig. 3) with temperature in the range $\sim 50 - 300$ K.

The cloud-free atmospheres have a high-pressure (100 bar) bottom boundary with zero albedo.

To assess the influence of thick clouds on the visibility of Raman features, we approximate the presence of a cloud deck at a given pressure by terminating the atmosphere and placing a Lambertian bottom boundary of our model atmosphere at that pressure level. Clouds can have a range of reflective/absorbing properties, which we incorporate by assigning different albedo values to the bottom boundary.

Additionally, we make a model atmosphere composed entirely of molecular nitrogen $^{14}N_2$, with a bottom boundary with albedo equal to 0.3 at a pressure level of 1-3 bar, and a TP profile matching that of the Earth's atmosphere[2]. Even though nitrogen-dominated atmospheres will most likely have other constituents that contribute to Rayleigh and Raman scattering (for example O_2 in the case of the Earth's atmosphere), we include only N_2 in order to investigate the effects of Raman scattering from this molecule alone. This analysis serves primarily as a demonstration that Raman scattering can be used to find signatures of gases other than H_2. Modeling realistic atmospheres of smaller planets is beyond the scope of this study. The mean molecular weight for the N_2 atmosphere is 28.0 m_p, and the surface gravity of the planet is $g = 10^3$ cm s^{-2}.

5.3.2 Radiative Transfer Calculations

We perform radiative transfer calculations using a discrete ordinate algorithm DISORT[3] (Stamnes et al. 1988). The code was designed to perform monochromatic unpolarized radiative transfer in a scattering, absorbing and emitting plane-parallel medium. The medium is divided into $N_{\rm lyr}$ homogeneous layers, separated by $N_{\rm lyr}+1$ levels. The user defines the optical depth (τ) and the single-scattering albedo (ω) for each layer. The medium can be illuminated by a parallel beam of radiation at the top boundary, at an angle of incidence specified by the user (polar angle $\Theta_{\rm in}$,

[2]We use the data from the *Planetary Atmospheres Node* `http://atmos.nmsu.edu/planetary_datasets/earth_temppres.html`.

[3]`ftp://climate1.gsfc.nasa.gov/wiscombe/Multiple_Scatt/`

azimuth angle $\Phi_{\rm in} = 0$). The code computes radiation intensities at user-selected angles and levels. All our calculations are performed with 50 computational levels ($N_{\rm lyr} = 49$) and 16 streams.

To define computational levels for each model atmosphere used in the radiative transfer calculations, we start by choosing the minimum and maximum pressure. For minimum we choose $P_{\rm min} = 10^{-4}$ bar. We tried a few different values of $P_{\rm min}$ (in the range $10^{-5} - 0.01$ bar) and the final output of the calculation did not change significantly. As mentioned before, the maximum pressure is $P_{\rm max} = 100$ bar for the clear atmospheres, 0.1-1 bar for the cloudy H_2/He, and 1-3 bar for the nitrogen-dominated atmospheres.

We divide the specified pressure range into computational pressure levels evenly distributed in logarithmic space. The temperature of each level is calculated using the temperature-pressure profile of the atmosphere. For each layer we compute its optical depth as

$$\tau_i = \frac{\sigma(T_i)}{\mu g} \Delta P_i \ , \tag{5.1}$$

where is μ the mean molecular mass, g is the surface gravity of the planet, and σ is the total scattering cross section (with contributions from Rayleigh and Raman scattering) calculated for the local temperature of that layer. The temperature of a layer is calculated as the average of the corresponding upper- and lower-level temperature, and the change in pressure across a layer ΔP_i is just the difference in pressure between the upper and the lower level.

Single-scattering albedo is computed for every computational layer and level using cross sections for Rayleigh and Raman scattering by H_2 and Rayleigh scattering by He atoms:

$$\omega(\lambda) = \frac{f_{\rm H_2}\sigma_{\rm H_2,Ray}(\lambda) + f_{\rm He}\sigma_{\rm He,Ray}(\lambda)}{f_{\rm H_2}\sigma_{\rm H_2,Ray}(\lambda) + f_{\rm He}\sigma_{\rm He,Ray}(\lambda) + f_{\rm H_2}\sigma_{\rm H_2,Ram}(\lambda)} \ , \tag{5.2}$$

where f denotes the number fraction of each species. In a pure N_2 atmosphere we use only $\sigma_{\rm N_2,Ray}$ and $\sigma_{\rm N_2,Ram}$. We take into account rotational and (ro-)vibrational Raman transitions from the ground electronic and ground vibrational level of the molecule. The initial rotational levels range from $J = 0$ to $J = 9$ for hydrogen, and from $J = 0$ to $J = 25$ for nitrogen. They are populated according to the Boltzmann distribution at the local temperature of the atmosphere T:

$$f_J = \frac{g_J}{Z} e^{-E_J/k_B T} \ , \tag{5.3}$$

where f_J is the fraction of molecules that are in a state with the rotational quantum number J, Z is the partition sum, E_J is the energy of the initial state,and g_J its statistical weight given by

$$g_J(\mathrm{H_2}) = \begin{cases} 2J+1 & \text{for } J \text{ even} \\ 3(2J+1) & \text{for } J \text{ odd} , \end{cases} \tag{5.4}$$

$$g_J(\mathrm{N_2}) = \begin{cases} 6(2J+1) & \text{for } J \text{ even} \\ 3(2J+1) & \text{for } J \text{ odd} . \end{cases} \tag{5.5}$$

The final molecular states are either in the ground or the first excited vibrational level. For more details on the cross sections used, see Appendix 5.A.

The publicly available version of the DISORT code (we use version 1.3 written in Fortran) does not include treatment of Raman scattering. We modify the code by including an additional source term due to Raman scattering (S_{Ram}) into the radiative transfer equation:

$$\mu \frac{dI_\lambda(\lambda, \hat{k})}{d\tau} = I_\lambda(\lambda, \hat{k}) - S_{\mathrm{Ram}}(\lambda) - \frac{\omega(\lambda)}{4\pi} \oint d\Omega P(\hat{k}, \hat{k}') I_\lambda(\lambda, \hat{k}') , \tag{5.6}$$

where $I_\lambda(\lambda, \hat{k})$ is the specific intensity of radiation per unit wavelength, at wavelength λ, propagating in the direction of $\hat{k}$, μ denotes the cosine of the emission angle, τ is the optical depth, and $P(\hat{k}, \hat{k}')$ is the (Rayleigh) scattering phase function.

The only source term included in the standard version of DISORT describes thermal emission for each layer of the atmosphere. Because we are focusing on the wavelength range in which thermal emission is not very important ($\lambda \lesssim 5000$ Å), we can modify this term to represent the Raman emission instead. As a result, our Raman source function is isotropic since it comes as a substitute for thermal emission, which was constructed with that property. The Raman emission is proportional to the mean radiation intensity at the Raman-shifted wavelength $J_\lambda(\lambda'_j)$, the Raman cross section for the transition in question $\sigma_{\mathrm{Ram}}(\lambda'_j)$ and the fractional population of the initial molecular state f_j:

$$S_{\mathrm{Ram}} = \sum_j J_\lambda(\lambda'_j) \frac{f_j \sigma_{\mathrm{Ram}}(\lambda'_j)}{\sigma_{\mathrm{tot}}(\lambda)} \left(\frac{\lambda'_j}{\lambda} \right)^3 , \tag{5.7}$$

The sum is taken over all relevant Raman transitions. The wavelength of the outgoing photon (λ) is related to the incident wavelength (λ'_j) through

$$\frac{1}{\lambda} = \frac{1}{\lambda'_j} - \Delta\nu_j , \tag{5.8}$$

where $\Delta\nu_j$ denotes the Raman shift in wavenumber (i.e. inverse wavelength) for the j-th transition. The factor $(\lambda'_j/\lambda)^3$ in the source term originates from the fact that Raman scattering conserves the number of photons, but redistributes them in wavelength space. In the radiative transfer equation we define the intensity in terms of energy per unit wavelength, hence we need to take that redistribution into account. One factor of λ'_j/λ comes from the change in the energy of the photons due to scattering and another $(\lambda'_j/\lambda)^2$ comes from the change in the wavelength range over which the photons get redistributed over (note that the Raman shift is constant in frequency, but changes in wavelength-space).

To calculate the Raman source term in each layer of the atmosphere at a given wavelength, we need to know the mean radiation intensity in that layer at Raman-shifted wavelengths, for all relevant Raman transitions. Therefore, we need to cover the entire spectral range wavelength by wavelength, record the mean intensities and then use them to compute Raman intensities at different wavelengths. We start this procedure at the shortest wavelength (3000 Å) and run the modified DISORT code without any Raman source terms. At each layer we record the mean intensity. After that, we run the code for longer wavelengths, up to 5000 Å with 1 Å spacing, now including the emission due to Raman scattering at shorter wavelengths. This accounts for Raman scattering transitions that lower the photon energy (i.e. the Stokes transitions). After we complete calculations over the entire wavelength range, we repeat them once again, this time having at our disposal the mean intensities over the entire wavelength range. This allows us to include the anti-Stokes Raman transitions (i.e. the Raman shifts to shorter wavelengths) as well.

The main output of the code is the radiation intensity at the top boundary in a specified direction (angle). To obtain the disk-integrated intensity we compute the outgoing intensity at several angles distributed across the planetary disk, as described in the next section and Appendix 5.B.

5.3.3 Geometric Albedo Calculation

Geometric albedo A_g is a measure of the reflectivity of a planet. It is defined as the ratio of the integral brightness of the planet at full phase to the brightness of a perfect Lambert disk of the same size and at the same distance from the star (see e.g. Hapke 2012). Geometric albedo can be calculated as the ratio of the total emergent intensity of the planet at zero phase angle $j(0)$ and the incident stellar flux πI_{inc} at

the planet-star distance (Sudarsky et al. 2000; Madhusudhan & Burrows 2012):

$$A_g = \frac{j(0)}{\pi I_{\rm inc}} \, . \tag{5.9}$$

Geometric albedo can be used to estimate the planet-star flux ratio observed on Earth as a function of the phase angle α (the angle between the line connecting the star and the planet and the line connecting the planet and the observer)

$$\frac{F_p}{F_*} = A_g \left(\frac{R_{\rm pl}}{a} \right)^2 \Phi(\alpha) \, , \tag{5.10}$$

where $R_{\rm pl}$ is the radius of the planet, a is the planet-star distance and $\Phi(\alpha) = j(\alpha)/j(0)$ is the phase function normalized at full phase.

In order to compute the total intensity of the planet $j(\alpha)$ at the phase angle α, we need to integrate the emergent specific intensity over the entire illuminated surface of the disk. To that end, we follow the procedure described in Horak (1950), Horak & Little (1965), and Madhusudhan & Burrows (2012). In the following equations $I(\eta, \zeta)$ denotes the specific intensity emerging from the part of the planetary surface with coordinates (co-latitude, longitude)=(η, ζ). Disk-integrated intensity is then given by

$$j(\alpha) = \int_0^{\pi} d\eta \sin^2 \eta \int_{\alpha - \pi/2}^{\pi} d\zeta I(\eta, \zeta) \cos \zeta \, . \tag{5.11}$$

The cosine of the angle of incidence and reflection is $\mu_0 = \sin(\eta)\cos(\zeta - \alpha)$ and $\mu = \sin(\eta)\cos(\zeta)$, respectively. These two angles are equal at full phase ($\alpha = 0$). To evaluate the integral in equation (5.11) numerically, first we make the following coordinate transformation:

$$\psi = \cos \eta \tag{5.12}$$

$$\nu = \sin \zeta \tag{5.13}$$

$$\xi = \frac{2}{\cos \alpha + 1} \nu + \frac{\cos \alpha - 1}{\cos \alpha + 1} \tag{5.14}$$

to get

$$j(\alpha) = \frac{\cos \alpha + 1}{2} \int_{-1}^{1} d\psi \sqrt{1 - \psi^2} \int_{-1}^{1} d\xi I(\psi, \xi) \, . \tag{5.15}$$

This double integral can be evaluated using standard quadrature methods (Horak & Little 1965)

$$j(\alpha) = \frac{\cos \alpha + 1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j I(\psi_i, \xi_j) \, , \tag{5.16}$$

where a_i, ψ_i are the Chebyshev weights and divisions, and b_j, ξ_j are the Gaussian weights and divisions. We use quadrature of order $n = 6$ with coefficients that are listed in Appendix 5.B.

Because of the symmetry of the $\alpha = 0$ problem, for $n = 6$ we only need to evaluate $I(\psi_i, \xi_j)$ at nine different points on the planet's disk. Hence, we run our radiative transfer code for nine different angles of incidence of the radiation beam, and record the emergent specific intensity at the same angle[4]. We sum up the results according to equation (5.16) to get the disk-integrated emergent intensity. We divide that with the incident irradiance to get the geometric albedo A_g.

5.4 Results

In order to test our implementation of Raman scattering in DISORT, we perform two tests to compare our results to those reported by Bétremieux & Yelle (1999) and Sromovsky (2005b). The results of these relatively simple test cases also nicely illustrate how Raman scattering affects the reflected light from a planet. Firstly, we look at the Raman-scattered light produced by a monochromatic source and then we illuminate the atmosphere with a flat photon spectrum with a single absorption line. After these simple test cases, we calculate the geometric albedo resulting from irradiating different model atmospheres with the solar spectrum (Kurucz 2005)[5].

5.4.1 Test Case I: Monochromatic Source

In this case, a H_2/He atmosphere is illuminated from the top with a beam of monochromatic light of wavelength $\lambda_{\text{beam}} = 1400$ Å. There is no incident radiation at other wavelengths. The bottom boundary is at the pressure of 10 bar and its albedo is set to zero. We record the total flux emerging from the top of the atmosphere. Figure 5.3 shows the emerging flux as a function of wavelength, normalized to the incident beam. The largest peak (at 1400 Å), accounting for $\sim 60\%$ of the reflected flux, corresponds to Rayleigh-scattered light, everything else is reflected light due to Raman scattering. The first group of peaks represents photons that are produced by pure rotational Raman transitions – anti-Stokes for the peaks at shorter wavelengths than the incident radiation and Stokes transitions for the peaks at longer wavelengths. The second group of peaks, starting around 1480 Å, and all consequent series correspond to combinations of vibrational and rotational transitions.

[4]Due to conventions used to define angles specifying the direction of radiation propagation, DISORT variables PHI0 and PHI which designate the azimuthal angles of the incident beam and the output, need to have a 180° offset in order to be oriented in the same direction.

[5]`http://kurucz.harvard.edu/sun/irradiance2005/irradthuwl.br_2nmresamp`

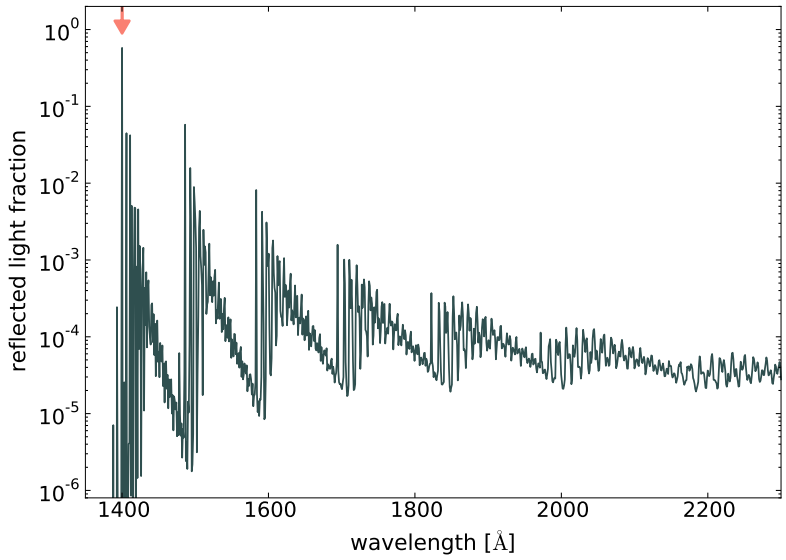

Figure 5.3: Reflected light for the case of monochromatic (λ = 1400 Å) radiation illuminating a clear 10 bar-deep H_2/He atmosphere. The arrow points at the largest peak which is the reflected light due to Rayleigh scattering at λ = 1400 Å. It accounts for about 60% of the total reflected light. Everything else comes from Raman scattering. The peaks around the incident wavelength are due to pure rotational Raman transitions. Peaks at wavelengths longer than ~1480 Å are due to various combinations of vibrational and rotational Raman transitions.

Bétremieux & Yelle (1999) and Sromovsky (2005b) perform similar tests of Raman scattering of monochromatic radiation. Our results shown in Figure 5.3 look very similar to results from both these papers, giving us confidence that our treatment of Raman scattering is sound. We do not expect our result to be in perfect agreement with theirs due to differences in the assumed atmospheric properties and taking into account Raman transitions from a different number of initial molecular states (we allow more excited rotational levels of the molecule to be initially populated, not just levels with $J = 0$ and $J = 1$).

5.4.2 Test Case II: Flat Spectrum with an Absorption Line

The second test case matches the one presented in Sromovsky (2005b). A 100-bar deep, cold-model (as defined in Section 5.3.1) atmosphere is illuminated by a beam of radiation with an absorption line at λ = 3100 Å inserted into an otherwise flat spectrum (constant λI_λ, with I_λ normalized at 2000 Å). Figure 5.4 shows the incident spectrum (dashed line) and the geometric albedo spectrum (solid line) calculated by our radiative transfer code. Our result for the hydrogen-dominated atmosphere

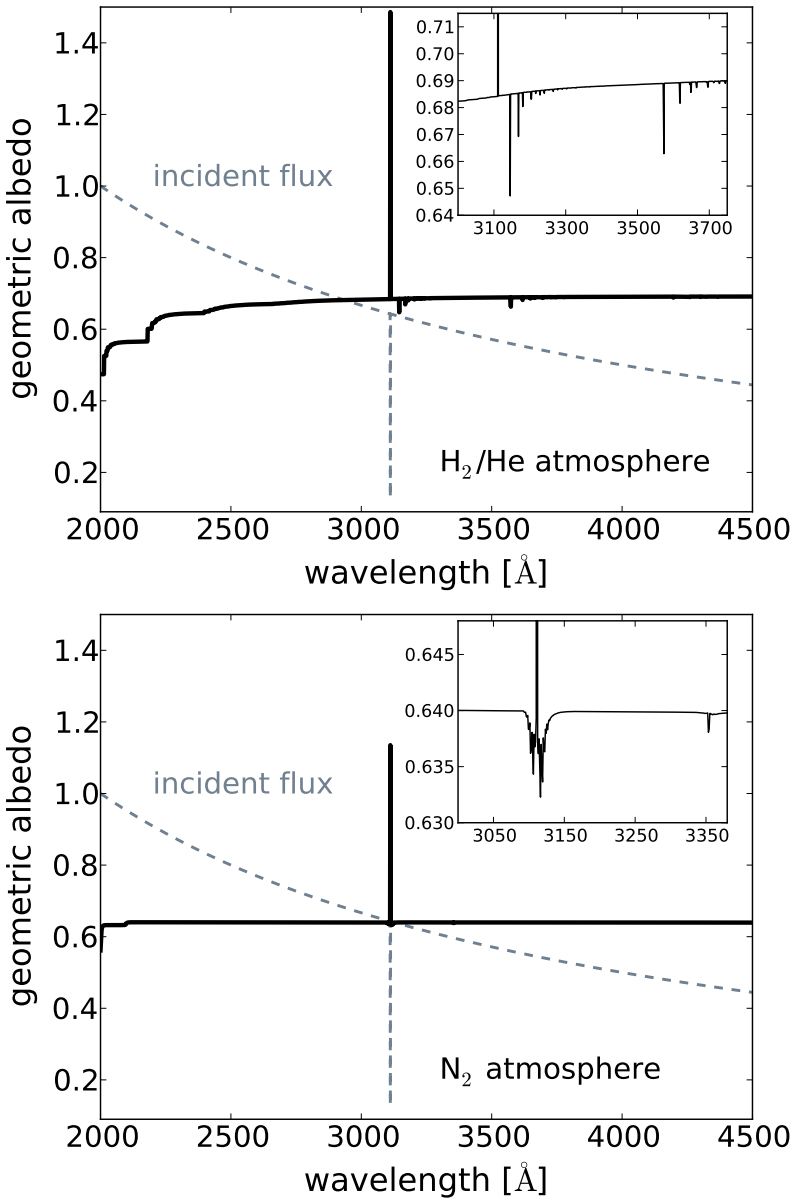

Figure 5.4: Upper panel: Solid line shows the geometric albedo for a clear, cold H_2/He atmosphere illuminated by a radiation beam with a flat photon spectrum with an absorption line at $\lambda \sim 3100$ Å (gray dashed line). The albedo is strongly enhanced at the wavelength of the absorption line. Raman ghost lines appear at Raman-shifted wavelengths (the inset shows a zoomed-in view of the ghost lines). Lower panel shows the same for a clear 100-bar atmosphere of pure N_2. The structure and positions of Raman ghost lines are different for the two molecules, H_2 and N_2, reflecting the structure of energy levels for each molecule. If observed, Raman ghost lines could be used to spectroscopically identify the composition of the scattering medium.

looks very similar to that of Sromovsky, although we cannot reproduce the exact values of the albedo because we do not include the effects of radiation polarization.

This test case demonstrates the main features of Raman scattering that were discussed in Section 5.2. At the position of the absorption line in the incident spectrum, there is a strong peak in the geometric albedo spectrum caused by photons that are Raman shifted into that wavelength. Dips in the geometric albedo – Raman ghosts of the absorption line – are less prominent, yet noticeable. Figure 5.4 shows that the positions of ghost lines in an H_2-dominated atmosphere are different from those in the atmosphere of N_2. Raman shifts are determined by the rotational and vibrational energy levels of the molecule and hence the relative shifts of ghost lines are unique for every species. This demonstrates how observations of Raman ghost lines could be used to spectroscopically identify homonuclear diatomics like H_2 and N_2 that otherwise do not have prominent spectral features in this wavelength range.

Intensities of individual ghost lines depend on the cross section for that particular transition and on the population of the initial state, governed by the Boltzmann distribution. By measuring the relative intensities of two ghost lines produced by the same feature in the incident spectrum, the temperature of the medium could be inferred.

5.4.3 Hydrogen Atmospheres Irradiated by Solar Spectrum

In Figure 5.5 we show the incident spectrum and the computed albedo spectra for three models of clear (cloud-free) H_2/He atmospheres. The three models have different temperature-pressure profiles, as described in Section 5.3.1. The albedo spectra show enhancements on the order of tens of percent at wavelengths at which the incident spectrum has absorption lines. In the cores of the most prominent lines, the Fraunhofer H and K lines of ionized calcium at 3969 and 3934 Å, the geometric albedo can be almost twice as high compared to the surrounding continuum.

The albedo spectra look very similar for all three cloud-free models, although there are some differences due to the temperature dependence of the population of molecular rotational levels. Raman transitions originating from different initial levels have (slightly) different cross sections, changing the overall strength of the Raman effect with temperature. The cross sections are highest for low-energy initial states (for example, see Figure 5.14 – the cross section for the rotational S-branch transition from level $J = 0$ is higher than the cross section for the initial level $J = 1$), making the Raman effect somewhat stronger in colder atmospheres.

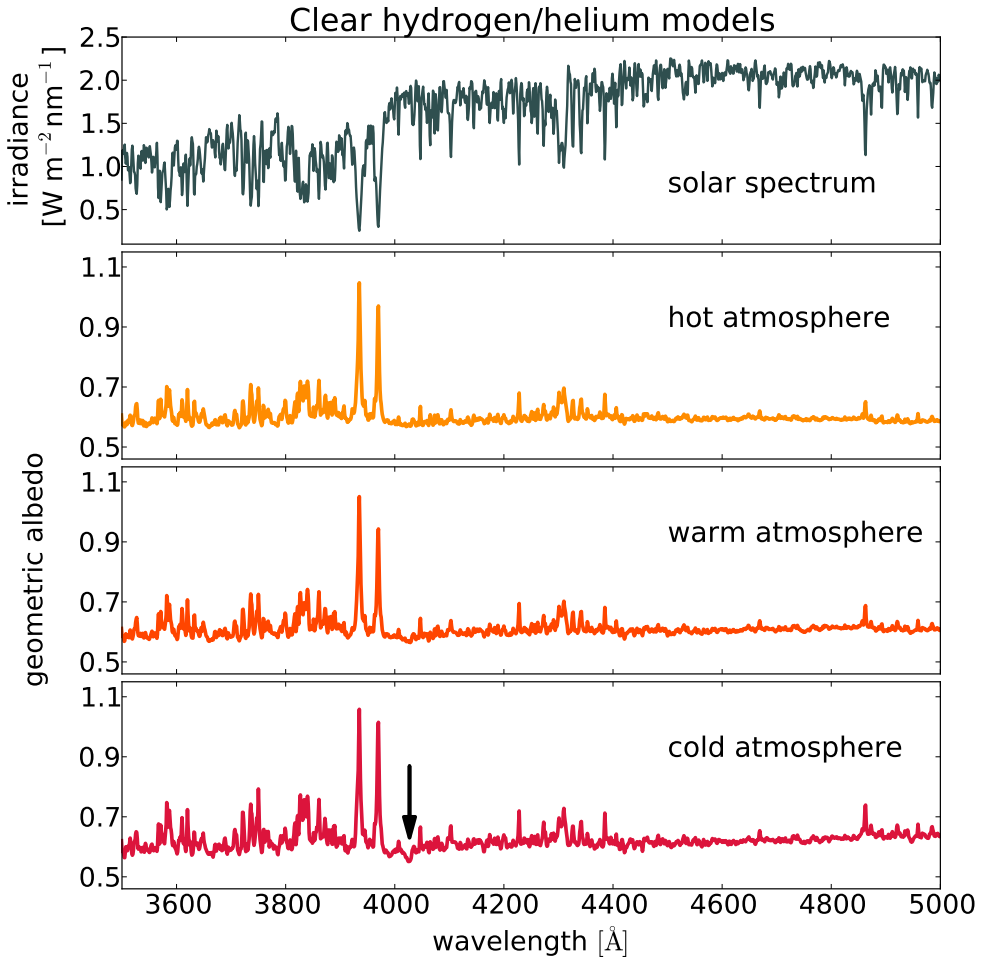

Figure 5.5: Three lower panels show the geometric albedo spectra calculated for clear 100-bar deep H_2/He atmospheres at different temperatures. The incident radiation has a solar-type spectrum (top panel). In an atmosphere with only Rayleigh scattering, the albedo spectrum would be flat, and thus all the features present in these spectra are due to Raman scattering. The most prominent Raman features are the albedo enhancements at wavelengths corresponding to strong absorption lines in the solar spectrum. The intensity of individual Raman ghost lines, such as the feature indicated by an arrow in the bottom panel, is much weaker and it is more significantly affected by the temperature of the atmosphere.

The position of ghost lines also depends on the initial state of the molecule, and hence the temperature of the medium. The most prominent ghost feature in the hydrogen spectrum is located at $\lambda \approx 4027$ Å (indicated by an arrow in the bottom panel of Figure 5.5). Two strong ghost lines nearly overlap at this wavelength – a ghost originating from an $S(0)$ Raman transition of the calcium H line, and the $S(1)$ transition from the calcium K line. In our cold atmosphere model, almost all hydrogen is in the $J = 0$ and $J = 1$ levels, maximizing the intensity of the ghost lines originating from these two states. In this model, the ghost feature at ~4027 Å causes a ~ 10% decrease in the albedo. At higher temperatures, the initial distribution of molecular states is broader, making individual ghost lines originating from low-energy states less intense.

Placing a thick cloud deck at different altitudes (i.e. different values of pressure) affects the overall albedo spectrum even more than changing the temperature, as shown in Figure 5.6. If a cloud deck is high in the atmosphere, the light passes through a smaller column of gas and experiences less scattering than when it travels through a clear atmosphere or an atmosphere with low-altitude clouds. Because the total column mass of gas above a certain level of the atmosphere is proportional to the ratio of the pressure at that level and the planet surface gravity, our results are degenerate with respect to these two quantities. Therefore, in order to measure the pressure level of clouds using the intensity of Raman peaks, the surface gravity of the planet needs to be known. We demonstrate this in Figure 5.6 where we show that the intensity of Raman features can be altered by changing the pressure level of the bottom boundary, while keeping the surface gravity fixed (left panel), or by changing the surface gravity while the bottom-boundary pressure remains the same (right panel).

What affects the strength of Raman features is the number of molecules available to photons for scattering above the cloud deck – having fewer Raman scattering events leads to an albedo spectrum in which Raman features are suppressed and the spectrum is dominated by Rayleigh scattering and the reflective properties of the clouds. Therefore, the strength of the Raman features can be used to infer the column mass of the scattering gas above the could deck and to put constrains on the presence and the altitude of clouds. In this optically thin regime, the albedo spectrum is much more sensitive to the properties of the planet than in the case of a clear, deep atmosphere where radiation can reach the pressure level at which the total optical depth (due to Rayleigh plus Raman scattering) becomes order unity. In

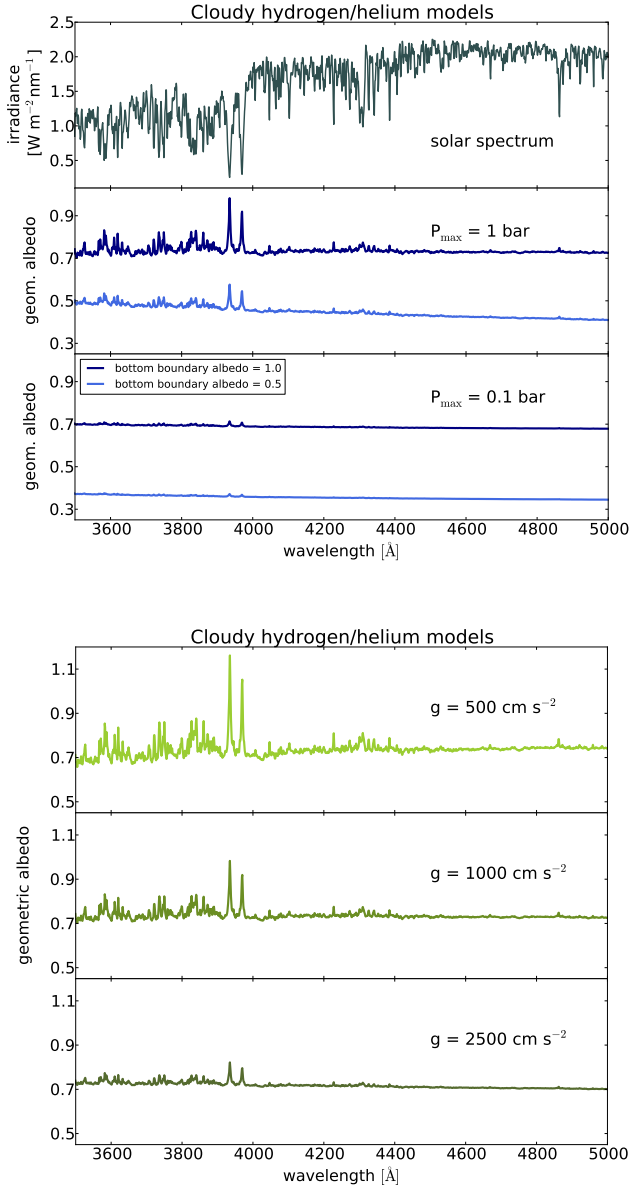

Figure 5.6: *Top:* Middle and bottom panels show geometric albedos for H_2/He atmospheres with a perfectly reflecting (albedo = 1.0) and partially absorbing (albedo = 0.5) bottom boundaries at the pressure value $P_{\rm max}$, representing a thick cloud deck. Spectral features due to Raman scattering are less prominent in shallow atmospheres because placing the cloud deck at higher altitudes (lower pressures) decreases the column of gas through which the radiation passes, thereby reducing the number of (Raman) scattering events. The column mass depends on the ratio of pressure and surface gravity, hence there is a degeneracy between these two quantities. The surface gravity of all these models is 10^3 cm s^{-2}. The incident solar spectrum is shown at the top. *Bottom:* Geometric albedo for H_2/He atmospheres with the bottom boundary at $P_{\rm max}$ = 1 bar with albedo of 1.0, for planets with different values of surface gravity.

the optically thin case, the overall planetary albedo depends on the albedo of the bottom boundary (i.e. the cloud deck) – more reflective bottom boundary results in a higher value of the geometric albedo continuum and larger Raman features as more photons get reflected from the bottom and have a second passage through the atmosphere on the way to the observer, allowing for more scattering evens to occur.

5.4.4 Nitrogen Atmospheres Irradiated by Solar Spectrum

We choose the parameters describing our nitrogen atmosphere so that they resemble the properties of Earth's atmosphere. We place the bottom boundary representing the surface of the planet at the pressure of 1 bar, and its albedo is set to 0.3. This choice of parameters leads to a relatively small column of the scattering gas, which is the reason why the albedo spectrum for this model shows only moderate effects of Raman scattering, as shown in Figure 5.7. The effects are stronger for deeper atmospheres, as shown for the atmosphere ending at the pressure of 3 bar. In the case of a small planet with a shallow atmosphere, Raman features alone are unable to give information on whether the bottom boundary is an opaque cloud deck or the planet surface.

5.5 Discussion

5.5.1 Probing Exoplanet Atmospheres with Raman Scattering

As shown in the previous section, the strength of Raman scattering features in the geometric albedo spectrum depends on the properties of the atmosphere. In this section, we discuss how observations of these features could be used to probe exoplanet atmospheres. We start from the most easily detectable Raman effects and move towards fainter features that require increasingly more precise observations of planetary albedo.

Constraining the Pressure Level of the Cloud Top/Surface

The most prominent and most easily detectable Raman features are the peaks in the geometric albedo at wavelengths corresponding to absorption lines in the incident spectrum. The albedo is enhanced at wavelengths corresponding to all lines, however the effect is most noticeable for strong lines, such as the Ca II H and K lines in the solar spectrum. At these wavelengths the albedo can be enhanced by almost a factor of ~ 2 in the case of a deep and clear H_2-rich atmosphere (Figure 5.5) or by tens of percent in atmospheres with thick clouds at pressures around 1 bar (Figure 5.6). In general, the strength of these features depends on the number of Raman scattering

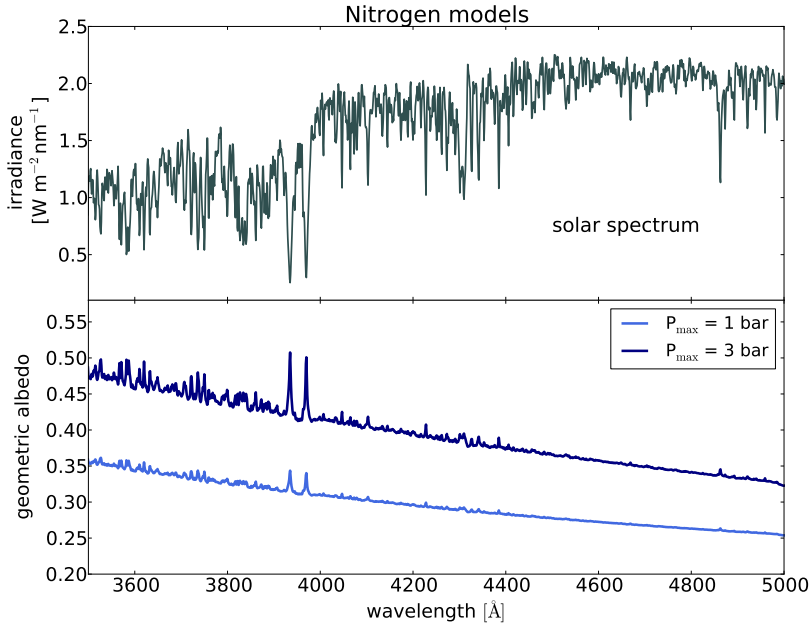

Figure 5.7: Bottom panel shows the geometric albedo for the model atmosphere composed entirely of N_2. The bottom boundary with albedo 0.3 is placed at the pressure $P_{\max}$. Higher pressure indicates a deeper atmosphere, leading to more Raman scattering events and thus producing stronger Raman features in the albedo spectrum. Top panel shows the incident spectrum.

events, i.e. on the column density of the scattering molecule that the light goes through before being scattered into our line-of-sight. This could be used to probe the depth of the atmosphere and put constraints on the pressure level of clouds or the planetary surface, if the surface gravity of the planet is known. If the dominant source of opacity at short wavelengths is scattering on particles (aerosols), instead of molecules, Raman features are not expected to appear in the spectrum.

Identifying the Dominant Scatterer

Raman scattering imprints 'ghosts' of features present in the incident spectrum at shifted wavelengths in the reflected spectrum. By measuring the Raman shift of the lines, the scattering molecule can be identified. As shown in Figure 5.4, the positions of Raman ghosts originating from the same line in the incident spectrum are different for H_2, N_2 and any other molecule. The intensity of individual ghost lines is small, on the order of a few percent of the albedo for the strongest lines. However, because Raman scattering occurs over the entire observed wavelength

range, small dips in the geometric albedo should appear at frequencies shifted by a fixed amount with respect to *all* stellar lines. Therefore, it might be possible to get a statistical detection of the ghost lines and to measure the Raman shifts by doing a cross-correlation analysis of the incident spectrum and the geometric albedo in frequency-space (Cochran 1981; Cochran et al. 1981).

In Figure 5.8, we show such a cross-correlation with our simulated spectra. The y-axis shows the fractional difference between the planet spectrum with Raman features included (F_{Raman}) and the planet spectrum with a flat albedo (F_{flat}) showing no signs of Raman scattering. The x-axis shows the wavenumber ($1/\lambda$) offset from a strong line in the incident spectrum. We select $\sim$ 5% of the wavelength bins in the analyzed wavelength range that correspond to the strongest absorption lines in the solar spectrum. We analyze how the planet spectrum around each of these strong lines is affected by Raman scattering, sum up their relative contributions, and present them in different colors for different model atmospheres.

At very small wavenumber shifts (i.e. close to the cores of stellar absorption lines) the planetary flux is enhanced by Raman scattering by a few tens of percent – the lines are 'filled-in'. In hydrogen-dominated atmospheres we see two troughs in the planet flux due to Raman scattering, around $\sim$ 354 cm^{-1} and $\sim$ 587 cm^{-1}. These flux decrements of the order of a few percent, correspond to Raman ghosts of strong stellar lines due to pure rotational Raman transitions from initial molecular states with $J = 0$ and $J = 1$. The troughs are strongest and most easily detectable for the cold atmosphere ($T \sim 100$ K) because these two molecular states account for the majority of H_2 molecules at such low temperatures. At higher temperatures more high-energy (high-J) states are populated and the intensity of individual ghosts is hence reduced. Raman shifts of molecular nitrogen are much smaller than those of hydrogen. Offsets corresponding to rotational Raman transitions from states with J in the range 0-12 are shown in Figure 5.8 by the gray hatched region. Because the distribution of Raman shifts in N_2 is so broad, it is more difficult to discern. Another difficulty associated with small Raman shifts is that ghost lines can overlap with line wings, especially for strong, broad lines like the Fraunhofer H and K lines. Raman ghosts can blend with parent lines and cause merely a decrease in the filling-in effect, instead of showing up as discrete troughs in the planetary flux, as shown in the case of the N_2 model in Figure 5.8.

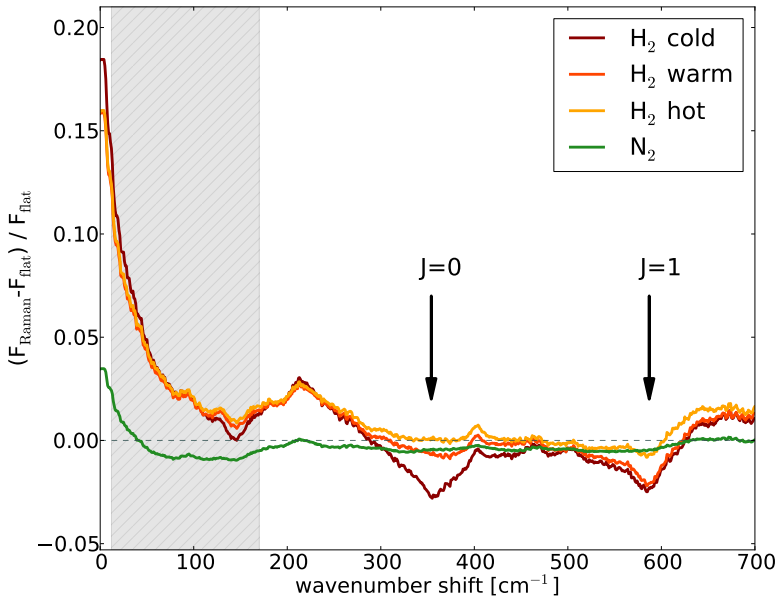

Figure 5.8: Cumulative deviation from a flat albedo spectrum in the vicinity of prominent absorption lines in the incident spectrum. At very small wavenumber separations the albedo is enhanced due to the filling-in effect of Raman scattering. Dips in the albedo are Raman ghost lines. By detecting the position of Raman ghosts in the albedo spectrum, the identity of the scattering molecule can be inferred. Arrows mark the expected positions of the H_2 ghost lines due to rotational Raman transitions from initial states with $J = 0$ and $J = 1$. Hatched region marks the expected positions of the N_2 ghosts for $J = 0, ..., 12$. The green curve shows a broad dip at these small wavenumber shifts and no discernible dips at the positions of H_2 ghost lines.

Constraining the Atmospheric Temperature

Raman transitions from different initial states of the molecule have different Raman shifts, which is why an individual absorption line in the incident spectrum produces a whole forest of ghost lines in the reflected light, as shown in Figure 5.4. The strength of a single ghost line depends on the total number of molecules in the corresponding initial state. For example, the intensity of a hydrogen ghost line resulting from a transition from the state with the rotational quantum number J is

$$I(J) \propto N_{H_2}(J) \, . \tag{5.17}$$

There is a degeneracy between the overall abundance of H_2 molecules and the fractional population of individual rotational states:

$$N_{H_2}(J) = N_{H_2}^{tot} f_J \ . \tag{5.18}$$

The fractional population of states is governed by the Boltzmann distribution, given by equation (5.3). From measuring the strength of just one ghost line $I(J)$, we cannot say whether that line is strong because the overall abundance of the molecule ($N_{H_2}^{tot}$) is high or because the fractional population of the corresponding initial state (f_J) if high. However, if we could measure at least two ghost lines and take their ratio, the total number of molecules cancels out and what if left is the relative population of these two initial states, which is related to the temperature of the atmosphere in the following way:

$$\frac{I(J_1)}{I(J_2)} \propto e^{-(E_{J_1} - E_{J_2})/k_B T} \ . \tag{5.19}$$

Because individual ghost lines are quite faint (up to a few percent of the albedo for the strongest lines) the uncertainties in the measured albedo should be well below a percent level to perform this type of measurement.

5.5.2 Detection Feasibility

In this section, we analyze the feasibility of observing Raman scattering in the atmospheres of nearby exoplanets, using the currently existing and the next generation observational facilities. First, we demonstrate that Raman features can be discerned in spectra of moderate spectral resolution.

Raman Scattering Features at Different Spectral Resolutions

Our reflection spectra were calculated on a uniform wavelength grid with 1 Å spacing. However, Raman features could be detected in observations performed with coarser spectral resolution than that. In Figure 5.9 we degrade our albedo spectra for the clear and cloudy (P_{max} = 1 bar) H_2/He atmospheres to spectral resolutions of R = 500, R = 200 and R = 50. Most Raman scattering features, even the strongest ones corresponding to calcium H and K lines in the solar spectrum, are smeared out in the low-resolution (R=50) spectrum. They can be more easily discerned in the medium-resolution (R=200) spectrum, especially if the atmosphere is clear. However, this model is probably very optimistic, and to identify these features in more realistic scenarios higher spectral resolutions (such as R = 500) would be preferred.

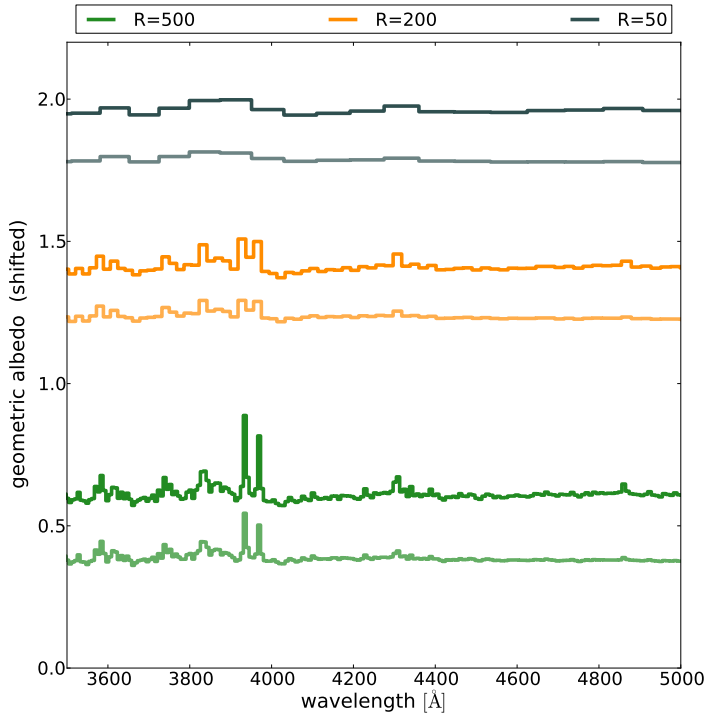

Figure 5.9: Calculated albedo spectra degraded to lower spectral resolutions of $R = 500$ (green), $R = 200$ (yellow) and $R = 50$ (gray). For every value of R, there is a pair of lines – the lighter line shows the albedo for a cloudy ($P_{\rm max} = 1$ bar) H_2/He atmosphere and the darker line is for a clear atmosphere. For ease of display, all albedos except the one shown in dark green (second from the bottom) are displaced on the y-axis.

Detecting Raman Features with High-contrast Instruments

In this section, we investigate the feasibility of observing Raman features using high-contrast imaging with a space telescope. Several approaches have been proposed, including both coronagraphs and starshades and using a range of telescope apertures. As an example in this initial study, we consider a 5 m telescope with a coronagraph achieving a contrast of 10^{-10}. We use the instrument noise model described by Robinson et al. (2016) to make our predictions. Here we give a brief overview of the main expressions used in their noise model and the baseline set of instrument parameters.

The photon count rate of a planet at a distance d from the observer is given by

$$c_p = \pi q f_{pa} \mathcal{T} \frac{\lambda^2}{hc\mathcal{R}} \left(\frac{D}{2}\right)^2 F_{p,\lambda}(d)\,, \tag{5.20}$$

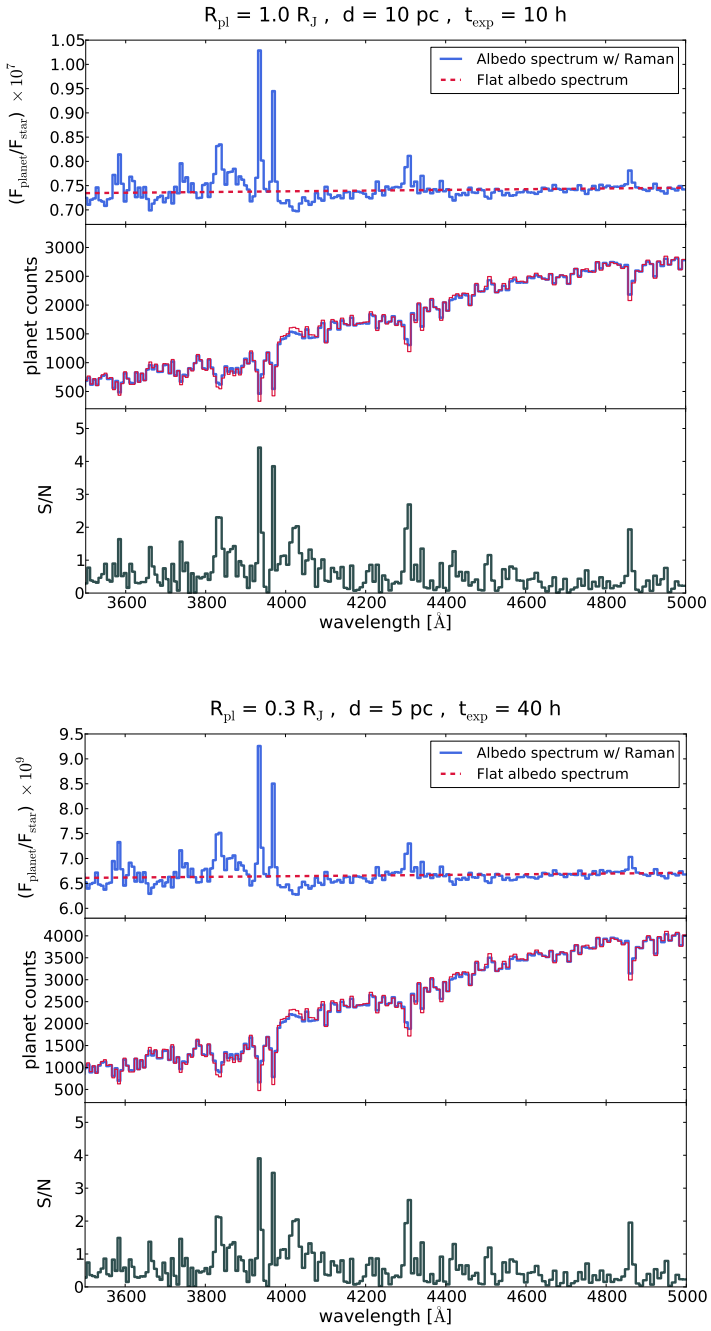

Figure 5.10: Top panel: the planet-star flux ratio (PSFR) for a Jupiter- and a Neptune-sized planet (left and right, respectively) orbiting a Sun-like star at 0.8 AU separation, observed at the phase angle of $\alpha = 80°$, at the distance of 10 pc and 5 pc, respectively. The blue line shows the PSFR from a model atmosphere that includes the effects of Raman scattering, degraded to spectral resolution of $R = 500$. The red dashed line shows a flat PSFR (i.e. without Raman features). Middle panel shows the total photon counts from the planet, observed with a 5-meter coronagraph-equipped telescope for the total exposure time of 10 and 40 hours. Bottom panel shows the ratio of signal to noise, at each spectral bin, to which signals with and without Raman features can be distinguished. Total S/N over all bins is ~ 17 in both cases.

Table 5.1: Astrophysical and telescope parameters.

Symbol	Description	Value
$R_{\rm pl}$	planet radius (R_J)	0.3 / 1.0
d	distance to observer (pc)	5 / 10
a	planet-star separation (AU)	0.8
α	phase angle (deg)	80
$\Delta t_{\rm exp}$	exposure time (hours)	40 / 10
q	quantum efficiency	0.9
f_{pa}	frac. of light in aperture	0.87
$\mathcal{T}$	throughput	0.05
C	coronagraph contrast	10^{-10}
$\mathcal{R}$	spectral resolution	500
D	telescope diameter (m)	5

where $F_{p,\lambda}(d)$ is the planet specific flux density, related to the stellar flux and the geometric albedo of the planet though equation (5.10). Descriptions and values of other parameters are listed in Table 5.1.

The background count rate consists of contributions from zodiacal (c_z) and exozodiacal (c_{ez}) light, stellar light leaked through the coronagraph (c_{lk}), dark current (c_D) and read noise (c_R):

$$c_b = c_z + c_{ez} + c_{lk} + c_D + c_R \, . \tag{5.21}$$

We use the expressions and parameters from Robinson et al. (2016) to compute all these different contributions to the background count rate. For the total exposure time $\Delta t_{\rm exp}$, the total number of background counts is $C_b = c_b \Delta t_{\rm exp}$ and the total number of photons coming from the planet is $C_p = c_p \Delta t_{\rm exp}$. The total noise count is $C_{noise} = \sqrt{C_p + 2C_b}$.

We estimate the signal-to-noise-ratio (S/N) to which Raman features in the albedo spectrum could be detected and distinguished from a flat, featureless albedo spectrum of an atmosphere in which there was no significant Raman scattering, due to an opaque high-altitude cloud deck. Our 'signal' then is the difference between the photon count coming from a planet whose albedo spectrum shows prominent Raman features ($C_{p,R}$, blue lines in Figure 5.10) and the photon count coming from a planet whose albedo spectrum would be featureless ($C_{p,f}$, red dashed lines in Figure 5.10). S/N is given by

$$S/N = \frac{|C_{p,R} - C_{p,f}|}{C_{noise}} = \frac{|c_{p,R} - c_{c,f}|}{\sqrt{c_{p,R} + 2c_b}} \sqrt{\Delta t_{exp}} \, . \tag{5.22}$$

In Figure 5.10 we show the planet-star flux ratio for a clear warm H_2/He atmosphere on a Jupiter- and Neptune-sized planets. Both planets are assumed to be orbiting Sun-like stars at a separation of 0.8 AU, observed at the phase angle of $\alpha = 80°$ (i.e. shortly before quadrature). Instead of just multiplying the calculated albedo with some specific planet phase function, we compute the disk integrated intensity at this phase by performing a new set of radiative transfer calculations that take into account different incident and reflected angles (μ_0 and μ) for $\alpha = 80°$, taken from Horak & Little (1965). The planets are assumed to be at a distance of 10 pc and 5 pc from Earth, for the larger and smaller planet respectively. We show the planet-star flux ratio observed on Earth and the photon counts assuming the planets are observed with a 5-meter coronagraph-equipped telescope, with spectral resolution $R = 500$ and for the total exposure time of 10 hours for the larger and 40 hours for the smaller planet. The bottom panel shows S/N to which we can distinguish between Raman features and a flat albedo spectrum, for each wavelength bin. As expected, the highest S/N is achieved in wavelength bins corresponding to strong Raman peaks associated with the Ca II H and K lines. The cumulative S/N for all wavelength bins is ~ 17 in both cases.

We perform the same type of analysis for a population of planets with radius $R = 1R_J$, located at a range of distances from Earth (from 5 pc to 20 pc) and for observations with a range of total exposure times (1-30 hours; all other telescope/observational parameters are as in Table 5.1). We show the total S/N over the wavelength range 3500-5000 Å in Figure 5.11. Our results suggest that detecting the signatures of Raman scattering in giant exoplanets could be feasible with a 5-meter coronagraph operating at these wavelengths. The analyzed distance range should contain several tens (or ~ 100) of solar-type stars.

Detection Method II: Blended Light Case

Next we consider observations in which the light from a planet is blended with the light from the host star. The principle is similar to observations using high-dispersion spectroscopy (e.g. Snellen et al. 2010; Brogi et al. 2012; Birkby et al. 2013; Lockwood et al. 2014). The total number of photons over a certain wavelength range reaching the observer on Earth from the planet-star system at the orbital phase Φ_i is given by $N_{s+p}(\Phi_i)$. At some other phase, Φ_j, the total number of photons is given by a data vector $N_{s+p}(\Phi_j)$. We first look at the covariance matrix between the

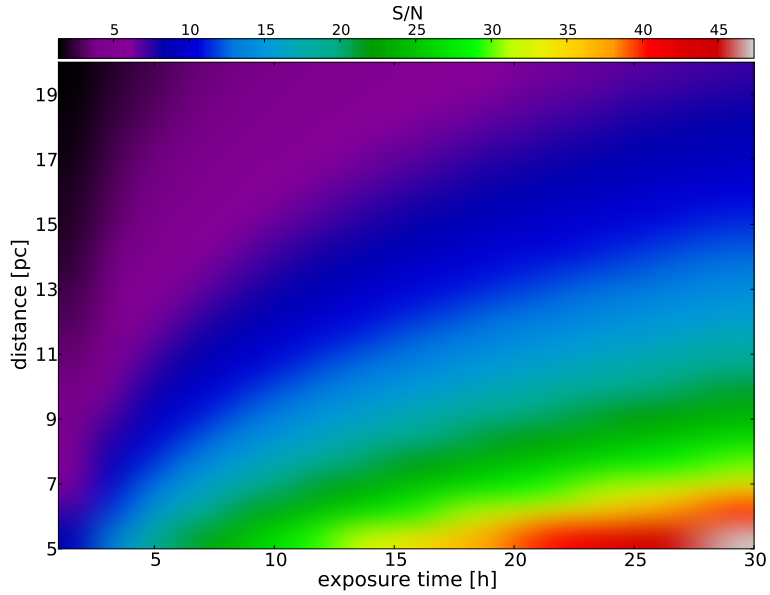

Figure 5.11: Cumulative S/N over the wavelength range 3500-5000 Å to which Raman features can be detected in a Jupiter-sized planet (orbiting the host star at 0.8 AU and observed near quadrature), depending on the distance to Earth and the total exposure time for observations with a 5-meter high-contrast (10^{-10}) telescope.

two vectors,

$$Cov(N_{s+p}(\Phi_i), N_{s+p}(\Phi_j)) \equiv C_{ij} = \delta_{ij} N_s \,, \tag{5.23}$$

where N_s is the number of photons coming from the star alone and δ_{ij} is the Kronecker symbol. The Fisher matrix is given by

$$F_{\alpha\beta} = \sum_{ij} [C^{-1}]_{ij} \frac{\partial N_{s+p}(\Phi_i)}{\partial p^\alpha} \frac{\partial N_{s+p}(\Phi_j)}{\partial p^\beta} \,, \tag{5.24}$$

where p denotes some set of parameters on which the covariance matrix does not depend. We find the inverse of the covariance matrix in equation (5.23) using the Taylor expansion

$$C_{ij}^{-1} = \frac{[N_{s+p}(\Phi_i) - N_{s+p}(\Phi_k)]^2}{2N_s} \,. \tag{5.25}$$

If we generalize this procedure to having observations at M different phases we get

$$\begin{aligned} C^{-1} &= \frac{1}{N_s}\left(1 - \frac{1}{M}\right) \sum_{i=1}^{M} N_{s+p}(\Phi_i)^2 \\ &- \frac{1}{N_s M} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j\neq i}}^{M} N_{s+p}(\Phi_i) N_{s+p}(\Phi_j) \,. \end{aligned} \tag{5.26}$$

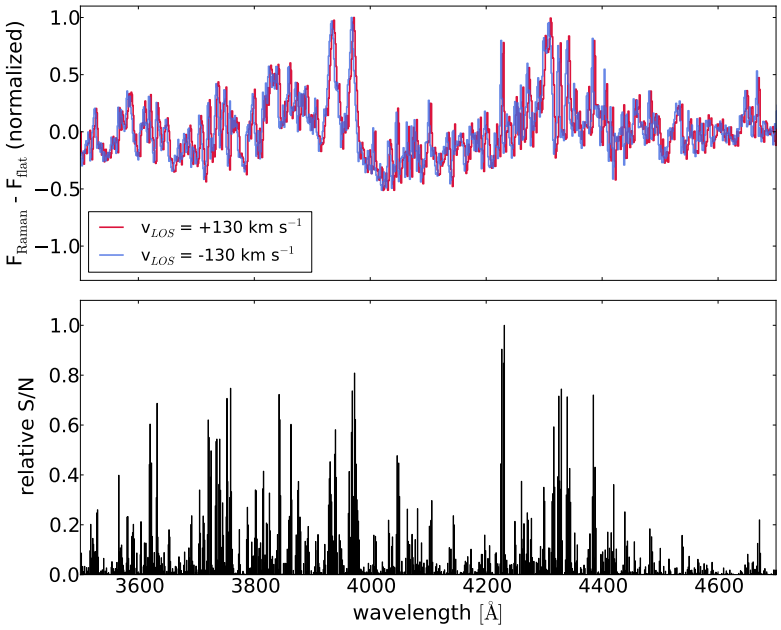

Figure 5.12: Upper panel: Raman features in the reflected light get blue(red)shifted as a function of the line-of-sight velocity v_{LOS} of the planet. Due to this property, the signal from the planet can be separated from the signal of its host star. This technique can be used to detect Raman scattering in the reflected light with high spectral resolution observations. The signal is dominated by wavelengths corresponding to strong and narrow lines, as shown by the normalized S/N distribution in the lower panel.

Taking the square root of this, we get the signal-noise ratio for each spectral bin:

$$S/N = \sqrt{C^{-1}} \, . \tag{5.27}$$

To get an estimate of how likely it would be to observe signatures of Raman scattering using this method, we use the simple two-phase formula (equation 5.25) for the case of a Jupiter-sized planet orbiting a Sun-like star at a separation of 0.05 AU. We assume that the planet is observed near both quadratures ($\alpha = \pm 80°$), where it has a radial velocity of ± 130 km s^{-1} with respect to the star, and we impart the blue(red)shift to the planetary spectrum accordingly. We define our signal as the difference between the observed photon number count in the case of a planetary atmosphere with Raman albedo features and an atmosphere with a flat, featureless albedo. This is shown in Figure 5.12 for the case of planet moving away from ($v = +130$ km s^{-1}) and moving towards the observer ($v = -130$ km s^{-1}). Raman

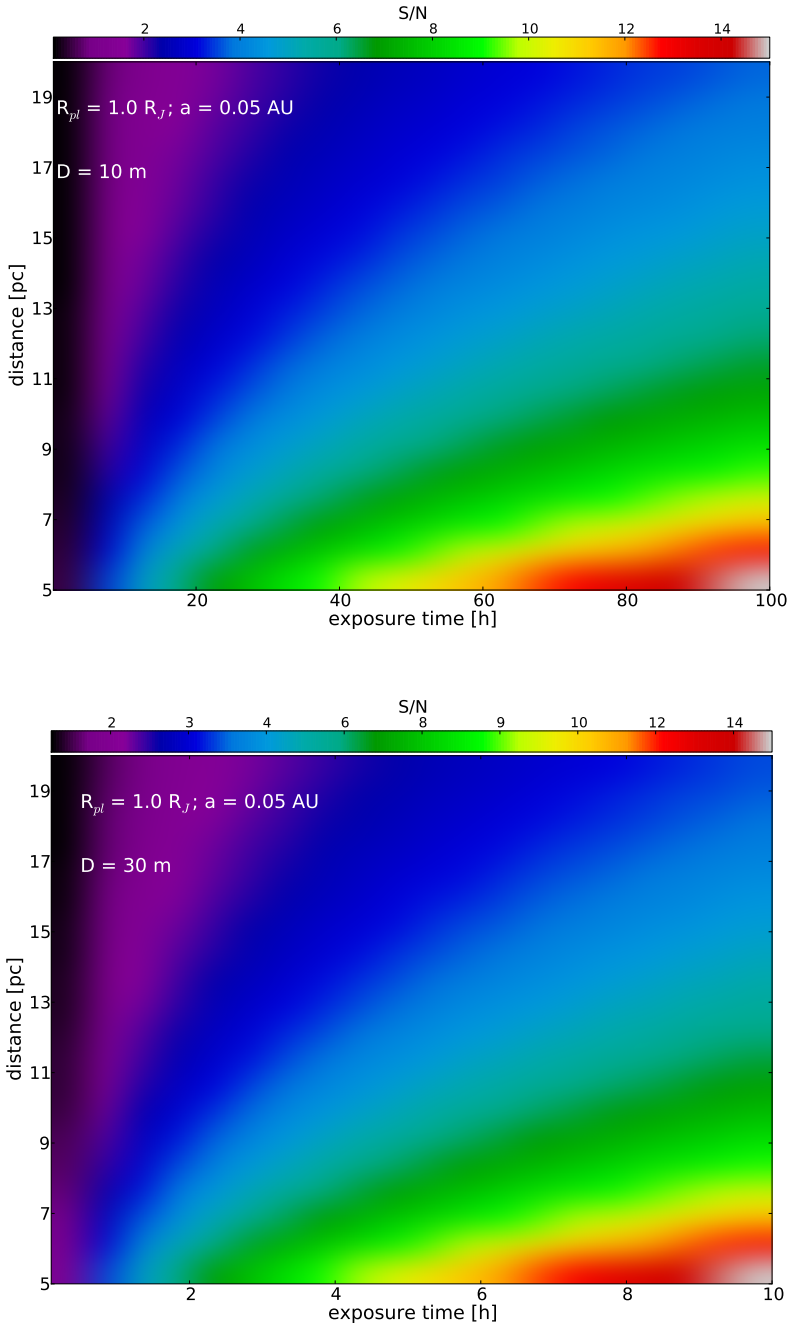

Figure 5.13: Estimates of the total S/N for observations of a Jupiter-sized planet orbiting a Sun-like star at 0.05 AU separation, observed near quadrature (α = 80°), for different combinations of distance to Earth and the total exposure time on a 10-meter (left panel) and a 30-meter ground-based telescope (right panel). Note the difference in the x-axis range in the two panels – the time needed to reach a certain S/N for the same planet on the smaller-aperture telescope is about an order of magnitude greater than the exposure time needed on the larger telescope.

features in the planetary spectrum get blue(red)shifted as the planet orbits the host star. The wavelength shift is only ~ 2 Å, hence moderately high spectral resolution is needed to detect these subtle changes with orbital phase. In this case we use $R = 3000$. The Doppler-shifting approach is more favorable for narrow spectral features, hence the contributions to the total S/N are more evenly distributed over the entire wavelength range than in the case of the previous method we considered, where most of the signal came from prominent (and fairly broad) Raman features associated with the stellar calcium lines.

In Figure 5.13 we show the calculated S/N using the two-phase formula, i.e. the difference between the red and the blue curve in Figure 5.12. The S/N is evaluated over the 3500 - 5000 Å wavelength range, for different combinations of distance from Earth and total exposure time on a 10-meter (left panel) or a 30-meter (right panel) ground-based telescope. Telescope throughput is assumed to be 15% over the analyzed wavelength range (in reality this value is not constant, it degrades towards the blue end of the spectrum). Our estimates suggest that detecting the signatures of Raman scattering in nearby exoplanets, using this method, would be very challenging with the currently existing telescopes. However, it should be possible with the next generation of large ground-based observatories if the photon-counting noise limit can be reached for this type of measurement.

5.5.3 Caveats and Future Work

In this study we focus our attention on atmospheric scattering, in particular on the effects of Raman scattering and its imprint in the geometric albedo of exoplanets. We do not include absorption from atomic or molecular species in our treatment of atmospheric radiative transfer. In an absorbing atmosphere the effects of scattering are less pronounced than in an atmosphere in which absorption is negligible. Absorption leads to an overall decrease in the geometric albedo, particularly at wavelengths corresponding to strong absorption lines, where albedo can be significantly diminished compared to out-of-line continuum. The most likely sources of absorption in the wavelength range that we analyze in this work ($\sim 300 - 500$ nm) are alkali atoms sodium and potassium, with absorption lines at 330 nm (Na) and 404 nm (K). The strength of these lines depends on the vertical profiles of the Na and K abundance in the form of free atoms. Atmospheric models used in this initial study are not complex enough to include these effects self-consistently.

Similarly to filling-in of *stellar* absorption lines by the Raman effect, the absorp-

tion lines from *atmospheric* absorbers also become less deep due to photons that are Raman-shifted into the lines. This effect could lead to underestimates in the inferred atmospheric abundances of absorbing species if Raman scattering is not properly treated. Sromovsky (2005b) found that this effect is present at a few percent level even in the near-IR, where Raman scattering is much weaker than at shorter wavelengths discussed in this work. According to their calculation, in a clear atmosphere of a Neptune-like planet, about 4% of the light coming from deep methane bands in the near-IR is Raman-scattered light. We leave the analysis of filling-in of atmospheric absorption lines in the near-UV and the visible part of the spectrum for future work.

5.6 Summary and Conclusions

We investigate the effects of Raman scattering by H_2 and N_2 molecules in the atmospheres of exoplanets. This process leaves distinct signatures imprinted in the reflected light and the planetary geometric albedo that could, if detected, be used to probe the physical conditions and the chemical composition of the atmosphere. Here we summarize the basics of Raman scattering, our results and conclusions:

- Raman scattering, unlike Rayleigh scattering, changes the frequency of the scattered photons. The intensity of Raman-scattered light at some frequency ν depends on the intensity of the incident light at the Raman-shifted frequency $\nu+\Delta\nu$. The frequency change $\Delta\nu$ is determined by the energy-level structure of the scattering molecule, and therefore can be used as a spectroscopic signature of the molecule.

- Raman-scattered photons fill-in absorption lines in the reflected spectrum that are caused by the lower radiation intensity at wavelengths corresponding to absorption lines in the incident spectrum. This causes sharp peaks in the albedo spectrum at the wavelengths of strong lines in the stellar spectrum, such as the calcium lines at $\lambda \sim 390$ nm in the solar spectrum. These albedo enhancements are the most prominent and most easily detectable Raman features in the albedo spectrum. Under favorable conditions of a deep and clear atmosphere, the albedo at these wavelengths can be enhanced by almost a factor of 2 compared to the surrounding continuum.

- Albedo enhancements are strongest for clear atmospheres and atmospheres with deep (low-altitude) clouds, whereas high-altitude clouds diminish the

effect of Raman scattering. Therefore, the strength of Raman peaks in the planetary albedo can be used to constrain the presence and the altitude of thick clouds.

- Raman scattering produces 'ghosts' of absorption lines in the incident spectrum. They are seen as small dips in the albedo at frequencies that are shifted by a fixed amount $\Delta\nu$ (Raman shift) from the original line in the stellar spectrum. Individual ghost lines are weak (on the order of a few percent change in the albedo), however they are present at fixed distances from all stellar lines and thus might be detected statistically using cross-correlation techniques. By detecting the ghost lines and measuring their Raman shifts, the scattering molecule can be uniquely identified. This might be the most direct way to *spectroscopically* confirm the presence of spectrally inactive molecules like H_2 or N_2 in the visible part of the spectrum.

- The strength of individual ghost lines depends on the population of the corresponding initial molecular state and hence depends on the temperature of the medium. If the intensity of ghost lines produced by the same stellar line, but originating from different initial molecular states, could be measured, the temperature of the atmosphere could be inferred.

- The most prominent Raman features can be discerned even in spectra of moderate spectral resolution of $R = \lambda/\Delta\lambda$ of a few hundred.

- Detecting Raman features in nearby hot exoplanets may be possible with the next generation observational facilities, such as the 30-meter-class ground-based telescopes.

- For exoplanets farther from their host stars, a promising method of detecting the signatures of Raman scattering would be using space-based coronagraph/starshade-equipped telescopes with spectrographs operating in the blue and near-UV wavelengths.

APPENDIX

5.A Rayleigh and Raman Scattering Cross Sections

5.A.1 Hydrogen

Many previous studies of Raman scattering used hydrogen Raman and Rayleigh cross sections from Ford & Browne (1973), which include transitions from initial molecular states with rotational quantum number $J \leq 3$. This is well suited for studies of cold atmospheres of giant planets in the Solar System. However, atmospheres of exoplanets can have much higher temperatures in which a significant fraction of molecular hydrogen populates states with $J > 3$. Since we were unable to find Rayleigh and Raman cross sections for $J > 3$ states in the literature, we calculate them using the formulae from Dalgarno & Williams (1962).

The cross sections defined in Dalgarno & Williams (1962) measure the the total (angle-integrated) outgoing power (in erg s^{-1}) divided by the incident energy flux (in erg cm^{-2} s^{-1}). We define cross section as the rate of scattering events (in units of events s^{-1}) divided by the incident photon flux (in photons cm^{-2} s^{-1}); thus we need to multiply the cross sections given in Dalgarno & Williams (1962) by the ratio of the outgoing (λ) and the incident (λ') wavelength. Because of this correction factor our cross sections have wavelength dependence proportional to $(\lambda'\lambda^3)^{-1}$, instead of $(\lambda)^{-4}$.

For a transition between the initial molecular state characterized by rotational and vibrational quantum numbers J_i and v_i, to the final state with quantum numbers J_f and v_f, the cross section is given by

$$Q_S(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3} \frac{(J_i + 1)(J_i + 2)}{(2J_i + 3)(2J_i + 1)} |\langle R_i|\gamma(r)|R_f\rangle|^2 \tag{5.28}$$

in the case of the S-branch transitions ($J_f = J_i + 2$);

$$Q_Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3} \left[3|\langle R_i|\alpha(r)|R_f\rangle|^2 + \frac{2}{3} \frac{J_i(J_i + 1)}{(2J_i - 1)(2J_i + 3)} |\langle R_i|\gamma(r)|R_f\rangle|^2\right] \tag{5.29}$$

for the Q-branch transitions ($J_f = J_i$; Rayleigh scattering is a special case of Q-branch transitions, one for which $v_i = v_f$ and $\lambda' = \lambda$); and

$$Q_O(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3} \frac{J_i(J_i - 1)}{(2J_i - 1)(2J_i + 1)} |\langle R_i|\gamma(r)|R_f\rangle|^2 \tag{5.30}$$

for the O-branch ($J_f = J_i - 2$). In the expressions above α and γ are the average polarizability and the polarizability anisotropy of the molecule, given as functions of

internuclear separation r. We use tabulated values of $\alpha(r)$ and $\gamma(r)$ from Rychlewski (1980). R represents the radial wave functions of the H_2 molecule for the initial and final ro-vibrational state (denoted by R_i and R_f, respectively) that are obtained by solving the Schrödinger equation using the H_2 potential from Wolniewicz (1993). Our computations are performed in the limit of static polarizability, and are thus valid for photon energies small compared to the first electronic excitation energy of the H_2 molecule.

The resulting cross sections are shown in Figure 5.14 for Rayleigh scattering and a few different Raman transitions. Table 5.A.1 contains coefficients that can be used to evaluate the Rayleigh and Raman scattering cross sections. The cross section for a transition from an initial molecular state characterized by quantum numbers $v = 0$ and J_i, to a final state with v_f and J_f is given by

$$Q(0, J_i; v_f, J_f; \lambda) = \frac{C}{\lambda' \lambda^3} \text{ [cm}^2\text{]}, \tag{5.31}$$

where λ and λ' are expressed in cm. The incident and scattered wavelengths are related by $\lambda'^{-1} = \lambda^{-1} + \Delta\nu$, where $\Delta\nu$ is the Raman shift of the transition in question.

Our values of cross sections for the most important transitions (Rayleigh, pure rotational Raman transitions and vibrational Q-branch transitions) agree with values of Ford & Browne (1973) to within $\sim 20\%$ at the shortest wavelengths (around 300 nm) and within $\sim 5\%$ around 500 nm. The source of discrepancy is most likely the fact that we do not include the effects of dynamic polarizability, which become more relevant at higher energies. The differences between our cross sections and the results of Ford & Browne (1973) for vibrational S- and O-branch transitions are much larger (between 50-100% difference). We were not able to establish the origin of this discrepancy. However, because these transitions are typically between one and two orders of magnitude weaker in cross section than the other Raman transitions (pure rotational and vibrational Q-branch), they do not significantly affect the main results of this work.

5.A.2 Nitrogen

For the case of molecular nitrogen, we use the approximate formulae from Dalgarno & Williams (1962) that rely on expanding the molecular polarizability and polarizability anisotropy around the equilibrium separation of nuclei r_0 according

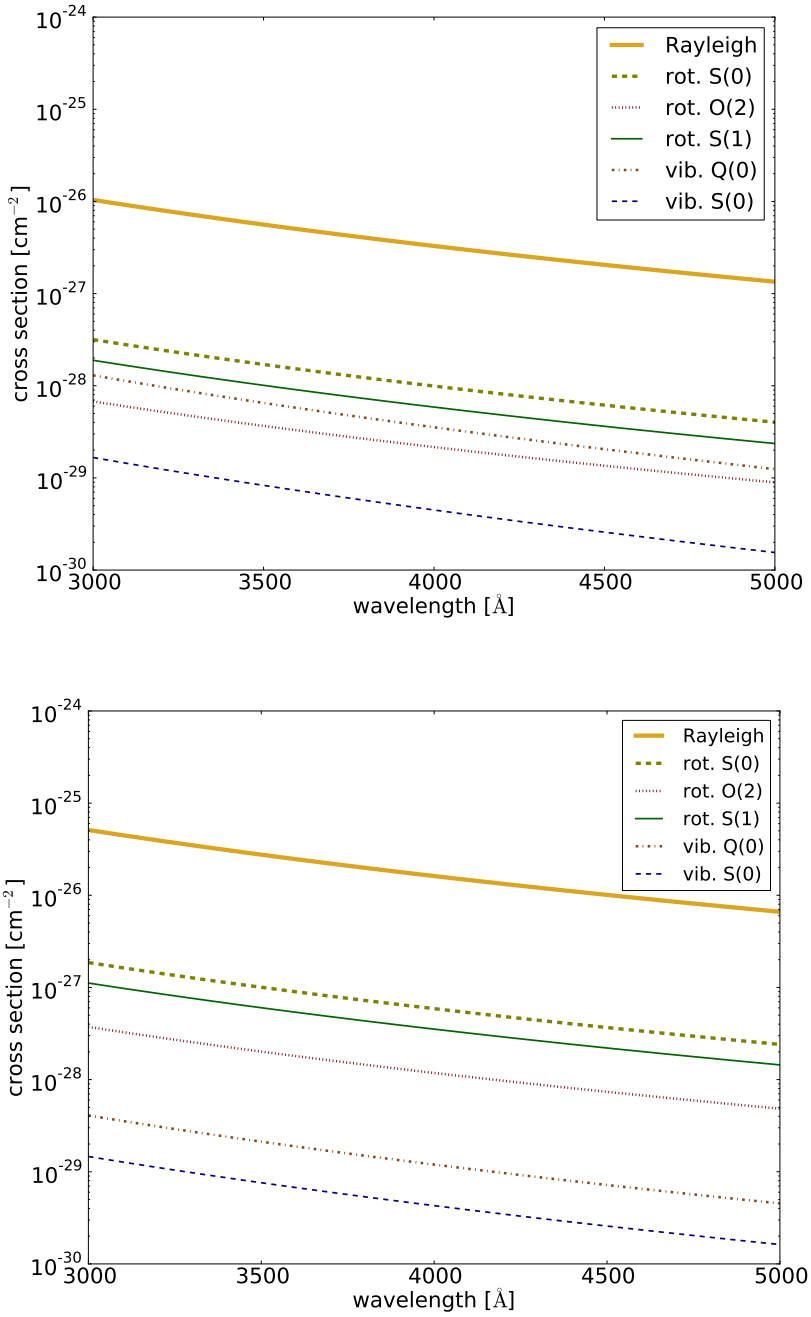

Figure 5.14: Cross sections for Rayleigh scattering and several Raman scattering processes, for H_2 (left panel) and N_2 (right panel) in the wavelength range relevant for this study. Designations of Raman transitions have the following meaning: 'vib' refers to vibrational transitions in which the vibrational quantum number of the molecule changes from 0 to 1 (ground to first excited state), and 'rot' describes pure rotational transitions in which the vibrational quantum number does not change. $S(i)$ denotes transitions in which the rotational quantum number increases by 2 (from i to $i+2$), for $Q(i)$ the rotational quantum number does not change, and for $O(i)$ the rotational quantum number decreases by 2.

J_i	J_f	v_f	C	$\Delta\nu[\mathrm{cm}^{-1}]$
0	0	0	$8.40817110301 \times 10^{-45}$	0.0
0	0	1	$1.56701934093 \times 10^{-46}$	4163.3085901
0	2	0	$2.64973960879 \times 10^{-46}$	354.5691479
0	2	1	$2.08231744526 \times 10^{-47}$	4500.1555355
1	1	0	$8.54368656301 \times 10^{-45}$	0.0
1	1	1	$1.66712462583 \times 10^{-46}$	4157.3872895
1	3	0	$1.61326313808 \times 10^{-46}$	587.3565875
1	3	1	$1.14413862330 \times 10^{-47}$	4715.3283129

Table 5.2: Coefficients for the scattering cross sections for H_2. The cross section for each transition is calculated according to equation (5.31). The column $\Delta\nu$ contains the value of the Raman shift for each transitions. The table is available in its entirety (including cross sections for initial states up to $J_i = 9$) in the machine-readable format on the journal website. This portion is shown here for guidance purposes.

to

$$\alpha(r) = \alpha(r_0) + (r - r_0)\alpha'(r_0) + ... \tag{5.32}$$

$$\gamma(r) = \gamma(r_0) + (r - r_0)\gamma'(r_0) + ... \tag{5.33}$$

and on approximating the vibrational wave function of the molecule by a harmonic oscillator eigenfunction with an eigenvalue $h\nu_e(v+\frac{1}{2})$. In the equations above, $\alpha'(r_0)$ and $\gamma'(r_0)$ are the derivatives of the polarizability and polarizability anisotropy with respect to internuclear separation, evaluated at the equilibrium distance.

The approximate formulae are:

- for Rayleigh scattering ($v_i = v_f$, $J_i = J_f$):

$$Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda^4}\left[3\alpha(r_0)^2 + \frac{2}{3}\frac{J_i(J_i+1)}{(2J_i-1)(2J_i+3)}\gamma(r_0)^2\right], \tag{5.34}$$

- for pure rotational S-branch transitions ($v_f = v_i$, $J_f = J_i + 2$):

$$Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\frac{(J_i+1)(J_i+2)}{(2J_i+3)(2J_i+1)}\gamma(r_0)^2, \tag{5.35}$$

- for pure rotational O-branch transitions ($v_f = v_i$, $J_f = J_i - 2$):

$$Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\frac{J_i(J_i-1)}{(2J_i-1)(2J_i+1)}\gamma(r_0)^2, \tag{5.36}$$

- for vibrational Q-branch transitions ($v_f = v_i + 1$, $J_f = J_i$):

$$Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\left[3\alpha'(r_0) + \frac{2}{3}\frac{J_i(J_i+1)}{(2J_i-1)(2J_i+3)}\gamma'(r_0)^2\right]\frac{(v_i+1)h}{8\pi^2\mu\nu_e}\,, \tag{5.37}$$

- for vibrational S-branch transitions ($v_f = v_i + 1$, $J_f = J_i + 2$):

$$Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\frac{(J_i+1)(J_i+2)}{(2J_i+3)(2J_i+1)}\gamma'(r_0)^2\frac{(v_i+1)h}{8\pi^2\mu\nu_e}\,, \tag{5.38}$$

- for vibrational O-branch transitions ($v_f = v_i + 1$, $J_f = J_i - 2$):

$$Q(v_i, J_i; v_f, J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\frac{J_i(J_i-1)}{(2J_i-1)(2J_i+1)}\gamma'(r_0)^2\frac{(v_i+1)h}{8\pi^2\mu\nu_e}\,, \tag{5.39}$$

where μ is the reduced mass of the molecule. For molecular nitrogen, we use the following set of parameter values: $\mu = 1.16 \times 10^{-23}$ g, $\nu_e = 6.988 \times 10^{13}$ Hz, $\alpha(r_0) = 1.778 \times 10^{-24}$ cm^3, $\gamma(r_0) = 7.14 \times 10^{-25}$ cm^3, $\alpha'(r_0) = 1.852 \times 10^{-16}$ cm^2, $\gamma'(r_0) = 2.185 \times 10^{-16}$ cm^2 (Bernath 2005; Pecul & Coriani 2002). Our cross sections agree to within $\lesssim 10\%$ with those used in studies of Raman scattering in the Earth's atmosphere (i.e. the Ring effect) by Chance & Spurr (1997).

5.A.3 Helium

For Rayleigh scattering on helium atoms we use the cross sections from Alipour et al. (2015).

5.B Quadrature Coefficients

To calculate the reflected light from a planet, we need to properly integrate over the surface of the planetary disk, as described in Section 5.3.3. In order to do that, we use the quadrature coefficients for order $n = 6$ and the corresponding angles of incidence for a planet at full phase (phase angle equal to zero) from Horak & Little (1965). Table 5.B gives the coefficients of the quadrature expansion (a and b), along with the coordinates (ψ and ξ) of the nine points on the disk used to compute the reflected light and the corresponding cosines of the incident angles (μ). At full phase, the contributions from all four quadrants of the disk are the same, and hence we can simply multiply the emerging intensity from the first quadrant by 4 to get the intensity from the entire planet.

number	latitude ($\pi/2 - \eta$)	longitude (ζ)	ψ	ξ	a	b	$\mu_0 = \mu$
1	0.2244 (12.86°)	0.2409 (13.81°)	0.222521	0.238619	0.426576	0.467914	0.94676541
2	0.2244 (12.86°)	0.7224 (41.39°)	0.222521	0.661209	0.426576	0.360762	0.73139233
3	0.2244 (12.86°)	1.2011 (68.82°)	0.222521	0.932470	0.426576	0.171324	0.35219154
4	0.6732 (38.57°)	0.2409 (13.81°)	0.623490	0.238619	0.274333	0.467914	0.75924690
5	0.6732 (38.57°)	0.7224 (41.39°)	0.623490	0.661209	0.274333	0.360762	0.58653110
6	0.6732 (38.57°)	1.2011 (68.82°)	0.623490	0.932470	0.274333	0.171324	0.28243569
7	1.1221 (64.29°)	0.2409 (13.81°)	0.900969	0.238619	0.0844887	0.467914	0.42135023
8	1.1221 (64.29°)	0.7224 (41.39°)	0.900969	0.661209	0.0844887	0.360762	0.32550020
9	1.1221 (64.29°)	1.2011 (68.82°)	0.900969	0.932470	0.0844887	0.171324	0.15673998

Table 5.3: The quadrature coordinates (latitude, longitude) or (ψ, ξ) and weights (a, b) used in calculating the integrated disk intensity according to equation (5.16). The last column (μ_0, μ) gives the cosine of the angle at which the radiation beam illuminates the atmosphere, and at which the light is reflected toward the observer (these two angles are equal at full phase).

References

Alipour, E., Sigurdson, K., & Hirata, C. M., 2015, PhRvD, 91.8, 083520, 083520

Belton, M. J. S., McElroy, M. B., & Price, M. J., 1971, ApJ, 164, 191

Belton, M. J. S., Wallace, L., & Price, M. J., 1973, ApJL, 184, L143

Bernath, P. F. 2005, Spectra of Atoms and Molecules

Bétremieux, Y., & Yelle, R. V., 1999, Icarus, 142, 324

Birkby, J. L., de Kok, R. J., Brogi, M., et al., 2013, MNRAS, 436, L35

Brinkmann, R. T., 1968, ApJ, 154, 1087

Brogi, M., Snellen, I. A. G., de Kok, R. J., et al., 2012, Nature, 486, 502

Brown, T. M., 2001, ApJ, 553, 1006

Cahoy, K. L., Marley, M. S., & Fortney, J. J., 2010, ApJ, 724, 189

Chance, K. V., & Spurr, R. J. D., 1997, ApOpt, 36, 5224

Cochran, W. D., 1981, Advances in Space Research, 1, 143

Cochran, W. D., & Trafton, L. M., 1978, ApJ, 219, 756

Cochran, W. D., Trafton, L., Macy, W., & Woodman, J. H., 1981, ApJ, 247, 734

Courtin, R., 1999, Planet. Space Sci., 47, 1077

Dalgarno, A., & Williams, D. A., 1962, MNRAS, 124, 313

Ford, A. L., & Browne, J. C., 1973, Atomic Data, 5, 305

Grainger, J. F., & Ring, J., 1962, Nature, 193, 762

Hapke, B. 2012, Theory of reflectance and emittance spectroscopy. (Cambridge, UK ; New York : Cambridge University Press, 2012)

Heng, K., Hayek, W., Pont, F., & Sing, D. K., 2012, MNRAS, 420, 20

Horak, H. G., 1950, ApJ, 112, 445

Horak, H. G., & Little, S. J., 1965, ApJS, 11, 373

Hubbard, W. B., Fortney, J. J., Lunine, J. I., et al., 2001, ApJ, 560, 413

Karkoschka, E., 1994, Icarus, 111, 174

Kurucz, R. L., 2005, Memorie della Societa Astronomica Italiana Supplementi, 8, 189

Lecavelier Des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D., 2008a, A&A, 481, L83

Lecavelier Des Etangs, A., Vidal-Madjar, A., Désert, J.-M., & Sing, D., 2008b, A&A, 485, 865

Lockwood, A. C., Johnson, J. A., Bender, C. F., et al., 2014, ApJL, 783, L29, L29

Long, D. A. 2002, The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules

Madhusudhan, N., & Burrows, A., 2012, ApJ, 747, 25, 25

Marley, M. S., Gelino, C., Stephens, D., Lunine, J. I., & Freedman, R., 1999, ApJ, 513, 879

Pecul, M., & Coriani, S., 2002, Chemical Physics Letters, 355, 327

Pollack, J. B., Rages, K., Baines, K. H., et al., 1986, Icarus, 65, 442

Price, M. J., 1977, Reviews of Geophysics and Space Physics, 15, 227

Robinson, T. D., Stapelfeldt, K. R., & Marley, M. S., 2016, PASP, 128.2, 025003

Rychlewski, J., 1980, Molecular Physics, 41, 833

Saumon, D., Marley, M. S., Abel, M., Frommhold, L., & Freedman, R. S., 2012, ApJ, 750, 74, 74

Schwieterman, E. W., Robinson, T. D., Meadows, V. S., Misra, A., & Domagal-Goldman, S., 2015, ApJ, 810, 57, 57

Seager, S., & Sasselov, D. D., 2000, ApJ, 537, 916

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S., 2010, Nature, 465, 1049

Sromovsky, L. A., 2005a, Icarus, 173, 284

—, 2005b, Icarus, 173, 254

Stamnes, K., Tsay, S.-C., Jayaweera, K., & Wiscombe, W., 1988, ApOpt, 27, 2502

Sudarsky, D., Burrows, A., & Pinto, P., 2000, ApJ, 538, 885

Sudarsky, D., Burrows, A., & Hubeny, I., 2003, ApJ, 588, 1121

Wallace, L., 1972, ApJ, 176, 249

Wolniewicz, L., 1993, JChPh., 99, 1851

Yelle, R. V., Doose, L. R., Tomasko, M. G., & Strobel, D. F., 1987, Geophys. Res. Lett., 14, 483

Chapter 6

HOW DOES THE SHAPE OF THE STELLAR SPECTRUM AFFECT THE GEOMETRIC ALBEDO OF EXOPLANETS AT SHORT OPTICAL WAVELENGTHS?

ABSTRACT

Diagnostic potential of the spectral signatures of Raman scattering, imprinted in planetary albedo spectra at short optical wavelengths, has been demonstrated in research on Solar System planets, and has recently been proposed as a probe of exoplanet atmospheres, complementary to albedo studies at longer wavelengths. Spectral features caused by Raman scattering offer insight into the properties of planetary atmospheres, such as the atmospheric depth, composition, and temperature, as well as the possibility of detecting and spectroscopically identifying spectrally inactive species, such as H_2 and N_2, in the visible wavelength range. Raman albedo features, however, depend on both the properties of the atmosphere and the shape of the incident stellar spectrum. Identical planetary atmospheres can produce very different albedo spectra depending on the spectral properties of the host star. Here we present a set of geometric albedo spectra calculated for atmospheres with H_2/He, N_2, and CO_2 composition, irradiated by different stellar types ranging from late A to late K stars. Prominent albedo features caused by Raman scattering appear at different wavelengths for different types of host stars. We investigate how absorption due to alkali elements, sodium and potassium, may affect the intensity of Raman features, and discuss the preferred observing strategies for detecting Raman features in future observations.

The material in this chapter has been submitted for publication to the AAS Journals.

6.1 Introduction

An important observable quantity in exoplanet observations is the flux received from the planet (F_p), measured relative to the flux of its host star (F_*). This ratio depends on the physical and orbital parameters of the planet-star system, such as the radius of the planet (R_p), its distance from the star (a), and the orbital phase angle (α) with respect to the observer, in the following way:

$$\frac{F_p}{F_*} = A_g \left(\frac{R_p}{a}\right)^2 \phi(\alpha) \, . \tag{6.1}$$

Most of the information about atmospheric properties is contained in the wavelength-dependent quantity A_g, called the geometric albedo. At short optical wavelengths, thermal radiation of planets is considered to be negligible compared to the reflected starlight. Therefore, by measuring the optical albedo of a planet, we are effectively measuring the reflective properties of its atmosphere and/or surface (e.g. Marley et al. 1999; Hu et al. 2012).

Measurements of the geometric albedo in the optical wavelength range have been obtained for a few dozen exoplanets to date (e.g. Rowe et al. 2008; Snellen et al. 2009; Christiansen et al. 2010; Désert et al. 2011; Kipping & Bakos 2011; Cowan & Agol 2011; Demory et al. 2013), mostly via photometric secondary eclipse observations with telescopes like *MOST*, *CoRoT*, and *Kepler*. Even though these broadband observations provide just one number for each planet—the value (or an upper limit) of the albedo averaged over the entire bandpass of the telescope—the albedos obtained so far span a broad range of values, pointing to a large diversity in the properties of exoplanet atmospheres. Most hot Jupiters have low values ($\lesssim 0.1$) of A_g (Heng & Demory 2013); however, there are some notable exceptions, such as *Kepler*-7b with $A_g = 0.35$ (Latham et al. 2010; Demory et al. 2011). Martins et al. (2015) observed the reflected light of a non-transiting hot Jupiter 51 Peg b in the wavelength range $\sim 380 - 690$ nm using ground-based high-resolution spectroscopy and found tentative evidence for a high average value of the geometric albedo[1]. Super-Earths observed with *Kepler* have been reported to have statistically larger geometric albedos compared to hot Jupiters (Demory 2014).

[1]There is an inherent degeneracy between the radius of the planet and its albedo (see Equation 6.1). The results of Martins et al. (2015) suggest geometric albedo values greater than unity for a radius of 1.2 R_{Jup}. The authors hence conclude that 51 Peg b may be an inflated hot Jupiter with a radius of 1.9 R_{Jup}, assuming $A_g = 0.5$. Birkby et al. (2017) present a more detailed discussion of that result, related to their recent mass measurement of 51 Peg b.

Evans et al. (2013) measured the albedo of hot Jupiter HD 189733 b over the wavelength range 290-570 nm using the *STIS* instrument on the *Hubble Space Telescope*. They report the geometric albedo binned in six wavelength channels, thus providing a low-resolution albedo spectrum of this exoplanet (shown in Figure 6.1). The measured albedo spectrum of HD 189733 b is broadly consistent with an atmosphere dominated by Rayleigh scattering and sodium absorption, as well as with models containing cloud particles of various compositions and sizes (Heng & Demory 2013; Barstow et al. 2014; Heng et al. 2014). The quality of the current data is insufficient to place tighter constraints on the sources of atmospheric opacity.

Nevertheless, albedo spectra of moderate spectral resolution ($R = 10^2 - 10^3$) at short optical wavelengths (3000 Å $\lesssim \lambda \lesssim$ 6000 Å) have the potential to provide valuable information about atmospheric properties, complementary to the information available at longer wavelengths (Burrows et al. 2008), which contain strong molecular absorption bands (e.g. Sudarsky et al. 2003; Cahoy et al. 2010; Morley et al. 2015). The opacity at short optical wavelengths is dominated by Rayleigh scattering on molecules or small particles (or by Mie scattering on larger particles, if they are present). Rayleigh scattering produces high values of albedo (a semi-infinite atmosphere with conservative Rayleigh scattering would have $A_g = 0.7975$; Prather 1974) and a smooth and featureless albedo spectrum (Marley et al. 1999; Sromovsky 2005a). However, Rayleigh scattering on molecules is always accompanied by Raman scattering, which does imprint spectral features over the smooth Rayleigh spectrum.

Raman features can be seen in the albedo spectra of many Solar System atmospheres, including the Earth's atmosphere. Figure 6.1 shows the albedo spectra[2] of giant planets and Titan obtained through ground-based observations by Karkoschka (1994), who identified between forty and sixty individual Raman features in each spectrum. The strongest features are visible in the spectra of the clear, Rayleigh-dominated atmospheres of Uranus and Neptune, in which $\sim 17\%$ of incident photons at 4000 Å get Raman scattered, whereas in aerosol-dominated atmospheres of Jupiter and Saturn this fraction is $\sim 5\%$ (Karkoschka 1998).

Building upon previous research on the Raman effect in the atmospheres of giant planets in the Solar System (e.g. Sromovsky 2005b; Bétremieux & Yelle 1999;

[2]All bodies except Jupiter are observed at small phase angles ($< 3°$ for Saturn and Titan; $< 1°$ for Uranus and Neptune), and hence the full-disk albedos are almost equal to the geometric albedos. The measured albedo of Jupiter is about 5% lower than its geometric albedo because the planet is observed at a phase angle of $\sim 10°$.

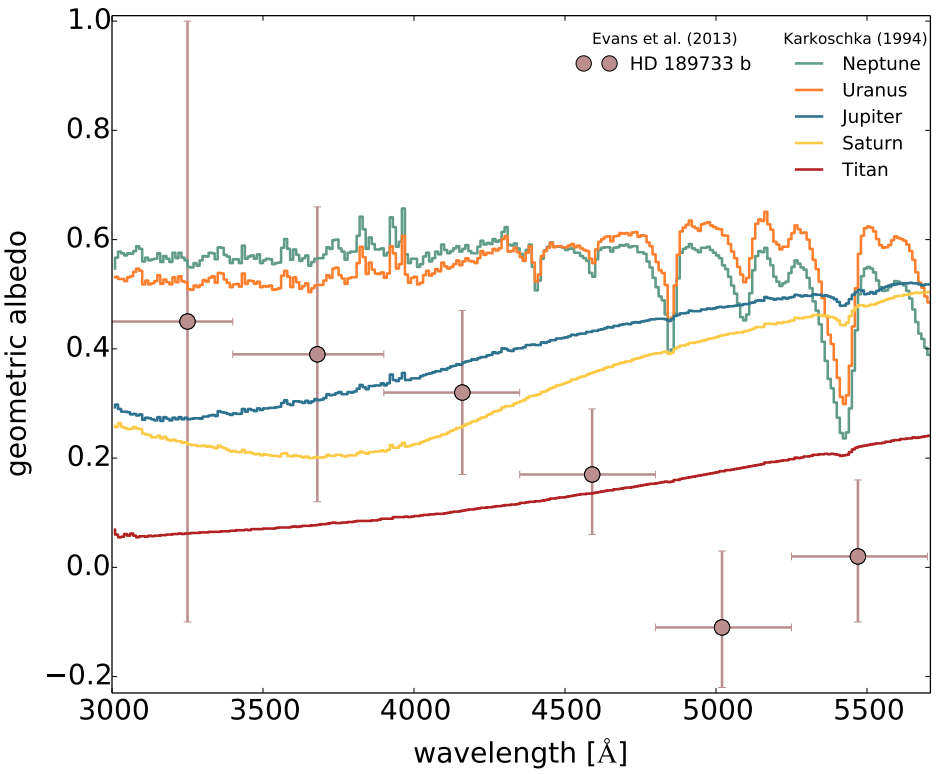

Figure 6.1: Lines show the measured full-disk albedo spectra of the jovian planets and Titan from Karkoschka (1994). The accuracy of the absolute albedo measurements is ~4%. The albedo spectra of Uranus and Neptune show prominent spectral features due to atmospheric absorption, as well as the Raman scattering features. The spectra of Titan, Jupiter, and Saturn are aerosol-dominated. The measured geometric albedo spectrum of hot Jupiter HD 189733 b from Evans et al. (2013) is shown for comparison.

Courtin 1999; Yelle et al. 1987; Wallace 1972; Belton et al. 1971; Belton et al. 1973; Price 1977; Cochran & Trafton 1978; Cochran 1981; Cochran et al. 1981), we recently demonstrated how spectral signatures of Raman scattering, imprinted in the albedo spectrum of gaseous exoplanets at short optical wavelengths, could be used to obtain information about atmospheric properties—composition, temperature, and the presence of clouds (Oklopčić et al. 2016, Chapter 5 in this thesis). Based on a simple feasibility analysis, we concluded that Raman features will likely be detectable in nearby exoplanets using the next generation of ground-based and spaceborne telescopes.

The intensity of Raman features depends on the properties of both the planetary

atmosphere and the incident stellar spectrum. Otherwise identical planetary atmospheres can produce a diverse range of albedo spectra depending on the spectral type of the host star, as we demonstrate in this chapter. In Chapter 5, we analyzed the effects of Raman scattering on the geometric albedo for different types of atmospheric models, all irradiated by the solar spectrum. In this chapter, we explore how Raman features in the albedo spectrum of an exoplanet change with changing the spectral type of the host star from late A to late K spectral types. By identifying the stars that can host planets with the most prominent Raman features over a given wavelength range, our goal is to help guide the selection of optimal targets for future observations of Raman scattering signatures in exoplanet atmospheres.

This chapter is structured as follows. In Section 6.2 we briefly introduce the physics of Raman scattering and its effects on the planetary geometric albedo spectra. In Section 6.3 we describe the input stellar spectra and the model atmospheres used in our radiative transfer calculations. In Section 6.4 we present the results in terms of the geometric albedo spectra of atmospheres of three different compositions (H_2/He, pure N_2, and pure CO_2) irradiated by five different stellar spectra. We investigate how the presence of atmospheric absorbers—alkali metals sodium and potassium—affects the albedo spectrum and the intensity of Raman features. Furthermore, we discuss the implications of our results in the context of choosing the optimal observing strategy and wavelength range for a given spectral type of planet-hosting stars. Finally, we summarize our conclusions in Section 6.5.

6.2 Raman Scattering

Raman scattering is an inelastic process in which the energy of the scattered photon changes due to its interaction with the molecular scatterer. Raman scattering is related to Rayleigh scattering, but characterized by much weaker cross sections (typically a few percent of the Rayleigh scattering cross section for the same molecule). Even though a minor fraction of the reflected light coming from a planet is due to Raman scattering, this component of radiation produces specific spectral features that carry information about atmospheric properties that the dominant (Rayleigh) component does not contain. The intensity of the Rayleigh component is directly proportional to the intensity of the incident light at the same wavelength, hence the Rayleigh-only albedo spectrum does not possess any sharp spectral features.

Due to inelastic Raman scattering, the rotational and/or vibrational state of the molecule changes and the energy difference is transferred to the scattered photon,

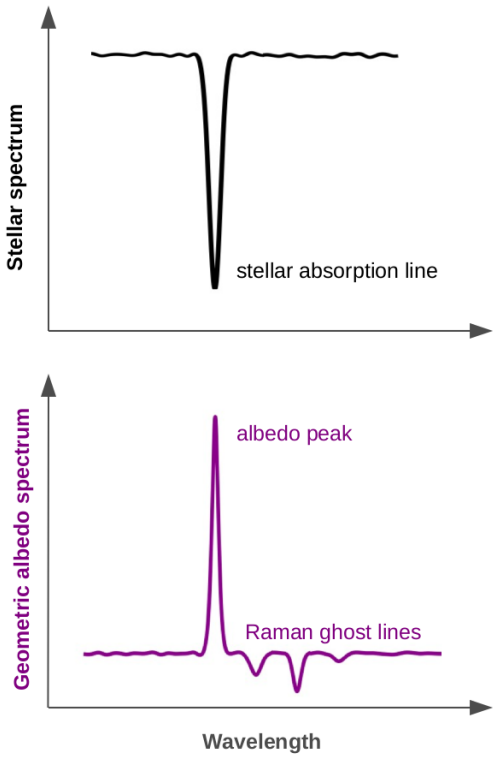

Figure 6.2: Two main types of Raman spectral features—the albedo peaks and the Raman ghost lines—produced by absorption lines in the incident stellar spectrum. Raman albedo peaks appear at the same wavelengths as the stellar absorption lines. Raman ghosts have a well-defined frequency shift with respect to the stellar lines that produce them.

which experiences a change in frequency. All spectral features present in the incident stellar spectrum, such as absorption or emission lines, get shifted in frequency in the Raman-scattered component of the reflected light. These frequency shifts result in two main types of Raman features present in the geometric albedo spectrum of a planet: (1) partial filling-in of absorption lines, resulting in sharp albedo peaks at the wavelengths corresponding to absorption lines in the stellar spectrum; and (2) a series of frequency-shifted absorption lines, resulting in small albedo decrements, called the Raman ghost lines (see Figure 6.2; for a more detailed description of how Raman features arise, see Chapter 5 and references therein).

Albedo peaks observed at high enough spectral resolution ($R \gtrsim 10^3$) can be quite prominent, especially at wavelengths of strong stellar absorption lines, such as the Ca II H and K lines in the solar spectrum. Raman ghosts are much fainter, producing changes in the albedo on the order of a few percent. The strength of the albedo peaks depends on the radiation-penetration depth of the atmosphere;

hence, by measuring their intensity we can obtain information about the presence and altitude of opaque clouds in the atmosphere. Raman ghost lines can be used to spectroscopically identify the molecule responsible for scattering and thereby infer the bulk atmospheric composition, even in the case of spectrally inactive, symmetric molecules which to not possess a permanent electric dipole moment, like H_2 and N_2. By comparing the intensity of individual ghost lines, information about the relative population of different molecular levels can be obtained and used to infer the temperature of the atmosphere.

6.3 Methods

6.3.1 Model Atmospheres and Radiative Transfer Calculations

In order to calculate the geometric albedo spectra, we perform radiative transfer calculations using models of planetary atmospheres. We construct three model atmospheres composed of:

1. hydrogen ($^1H_2$, 90% by number) and helium (^{4}He, 10% by number), with the mean molecular weight of $\mu = 2.3\ m_H$, where m_H is the mass of the hydrogen atom,
2. nitrogen, $^{14}N_2$, with $\mu = 28.0\ m_H$,
3. carbon dioxide, $^{12}C\ ^{16}O_2$, with $\mu = 44.0\ m_H$.

The surface gravity of the planet in all three cases is set to 1000 cm s^{-2}.

We assume the temperature-pressure profile that closely resembles the results of Cahoy et al. (2010) obtained for a warm Neptune-like planet at 0.8 au from the host (Sun-like) star, with the atmospheric temperature in the range $\sim 300 - 1000$ K. We do not compute the temperature profile for each atmosphere in a self-consistent way because the results of Chapter 5 show that the atmospheric temperature has only a small effect on the intensity of the most prominent Raman features—the albedo peaks—which are the main focus of this investigation. Atmospheric temperature has a much stronger effect on the structure of the less prominent Raman features, the Raman ghosts; however, we do not study them in detail in this chapter.

We analyze the propagation of radiation at short optical wavelengths ($\sim 3000 - 5500$ Å), for which scattering is expected to be the dominant source of opacity, and ignore absorption (except when otherwise noted) as well as thermal emission of the planet. The wavelength-dependent opacity and single scattering albedo are

Table 6.1: Names and properties of stars used as input spectra (from Valdes et al. 2004).

Name	Spectral type	$T_{\rm eff}$ (K)	$\log_{10}g$	[Fe/H]
HD 76644	A7V	-	-	-
HD 33256	F2V	6442	4.05	-0.30
HD 115617	G5V	5590	4.23	-0.03
HD 149661	K2V	5362	4.56	0.01
HD 147379	K7V	3720	4.67	0.00

calculated using cross sections for Rayleigh and Raman scattering of all the relevant species. The calculations used to obtain cross sections for Rayleigh and Raman scattering on hydrogen and nitrogen are described in Chapter 5. Additionally, in this paper we include an analysis of Raman scattering in an atmosphere composed of CO_2, which has not been studied before in the context of exoplanets. The computation of the relevant cross sections for CO_2 is described in Appendix 6.A. In Section 6.4, we discuss how the presence of alkali metals in the atmosphere may affect the detectability of Raman features. For the albedo spectra presented there, we included absorption by sodium and potassium as an additional source of opacity. The absorption cross sections used for that analysis are described in Appendix 6.B.

The plane-parallel atmospheres are divided into 49 homogeneous layers, evenly distributed in log-pressure from 10^{-4} to 100 bar. The atmospheric conditions (pressure, temperature, opacity, single-scattering albedo) are uniform within each layer and change discontinuously at the layer boundaries called levels. The bottom level is a surface with albedo set to zero. The atmosphere is illuminated at the top boundary by a beam of radiation incident at a specified angle. We perform radiative transfer calculations using a discrete ordinate algorithm DISORT (Stamnes et al. 1988), which we modified to include Raman scattering (see Chapter 5 for details). The code calculates the outgoing flux at the top boundary for every wavelength and the specified outgoing angle, which varies at different points on the planetary disk. We find the intensity for nine locations on the disk and use them to compute the total emergent intensity at zero phase angle using the Gaussian quadrature integration method of order $n = 6$ (Horak 1950; Horak & Little 1965; Madhusudhan & Burrows 2012). The geometric albedo is then calculated at each wavelength as the ratio of the total emergent intensity of the planet and the incident stellar flux.

6.3.2 Stellar Spectra

The structure and the intensity of Raman features strongly depend on the intensity of individual stellar lines and the overall shape of the stellar continuum. We use five different input spectra, taken from the Indo-US Library of Coudé Feed Stellar Spectra (Valdes et al. 2004). These spectra cover the wavelength range between 3460 Å and 9464 Å, with spectral resolution of $\sim$ 1 Å. We perform Gaussian smoothing and re-sampling to a uniform wavelength grid between 3500 Å and 5500 Å, with 2 Å spacing.

In order to properly model redistribution of photons in wavelength-space due to Raman scattering, we need to begin our radiative transfer calculation at shorter wavelengths than the wavelength range that we are interested in observing. For the strongest vibrational Raman transitions of molecular hydrogen, Raman frequency shifts range between $\sim$ 4200 and $\sim$ 4800 cm^{-1}. Therefore, we need to start our radiative transfer at $\lambda \approx 3000$ Å in order to properly treat the filling-in of lines at $\lambda \gtrsim 3500$ Å. Since the Valdes et al. (2004) spectra do not cover wavelengths that far into the blue, we extrapolated their spectra with those from the Pickles atlas (Pickles 1998) in the wavelength range 3000-3500 Å. We used the corresponding spectral type for each star, except in the case of A7V, which does not exist in the Pickles atlas, so we used the F0V Pickles spectrum instead. The Pickles spectra have a lower spectral resolution than the Valdes et al. spectra, as they are reported on a wavelength grid with $\Delta\lambda = 5$ Å. This should not significantly affect the filling-in of lines longward of 3500 Å, but it can cause the strength of Raman features in the low-resolution part of the spectrum to be underestimated. We report our main results—the geometric albedo spectra of planets around different spectral types of host stars—only for wavelengths $\lambda \geq 3500$ Å, for which we have stellar spectra of higher resolution and for which the filling-in of lines by photons originating from shorter wavelengths is properly treated.

The used stellar spectra (normalized at 5000 Å) are shown on the left-hand side of Figure 6.3. The selected stars span the range of spectral types from A7V to K7V. Table 6.1 lists the names of the five stars whose spectra are used, along with their spectral type, effective temperature, surface gravity, and [Fe/H].

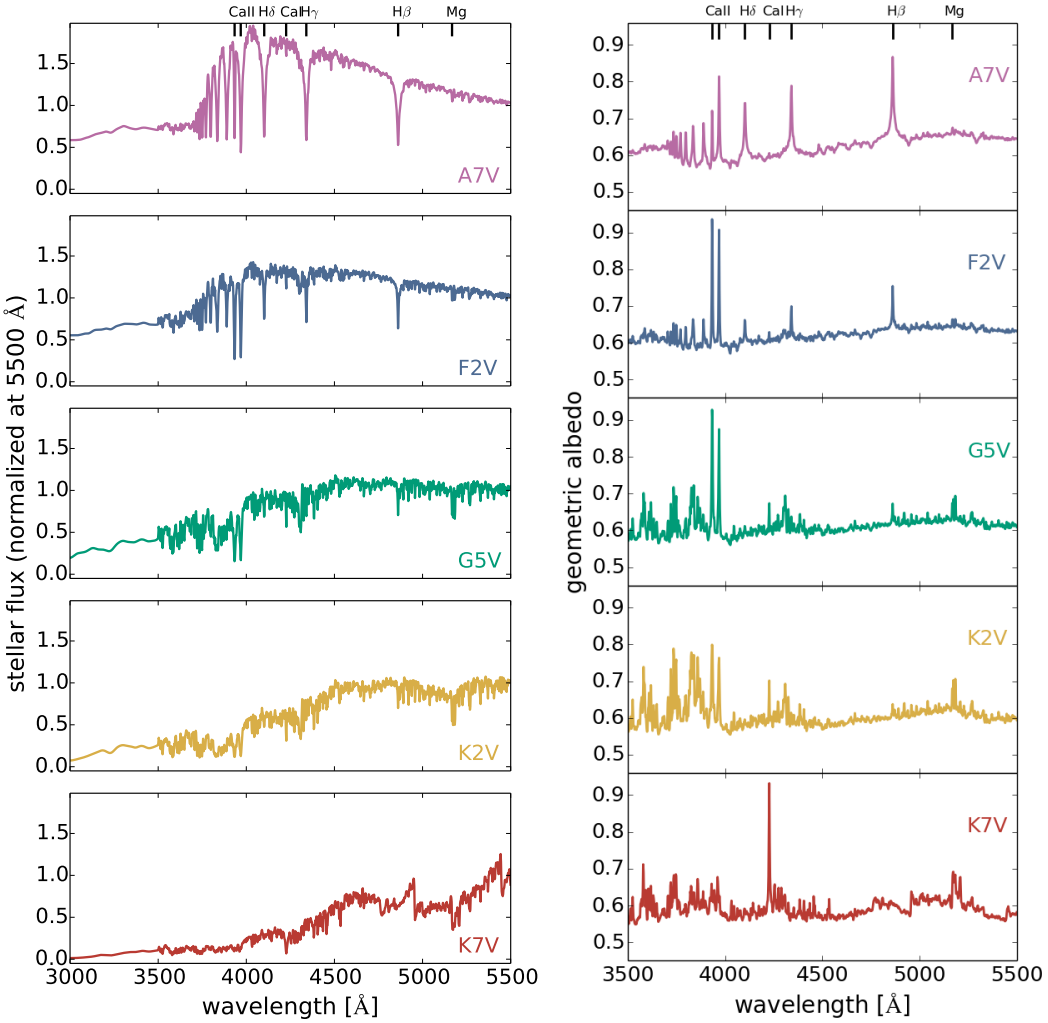

Figure 6.3: Left: stellar spectra of five stars from the Indo-US stellar library (Valdes et al. 2004) catalog. The stars' names and properties are given in Table 6.1. Bars on the top mark the positions of prominent spectral lines. Right: geometric albedo spectra calculated for clear, 100 bar deep hydrogen/helium atmospheres irradiated by the stellar spectra on the left-hand side. Albedo spectra have peaks at wavelengths corresponding to prominent stellar lines (some of which are indicated on the top) due to the effects of Raman scattering.

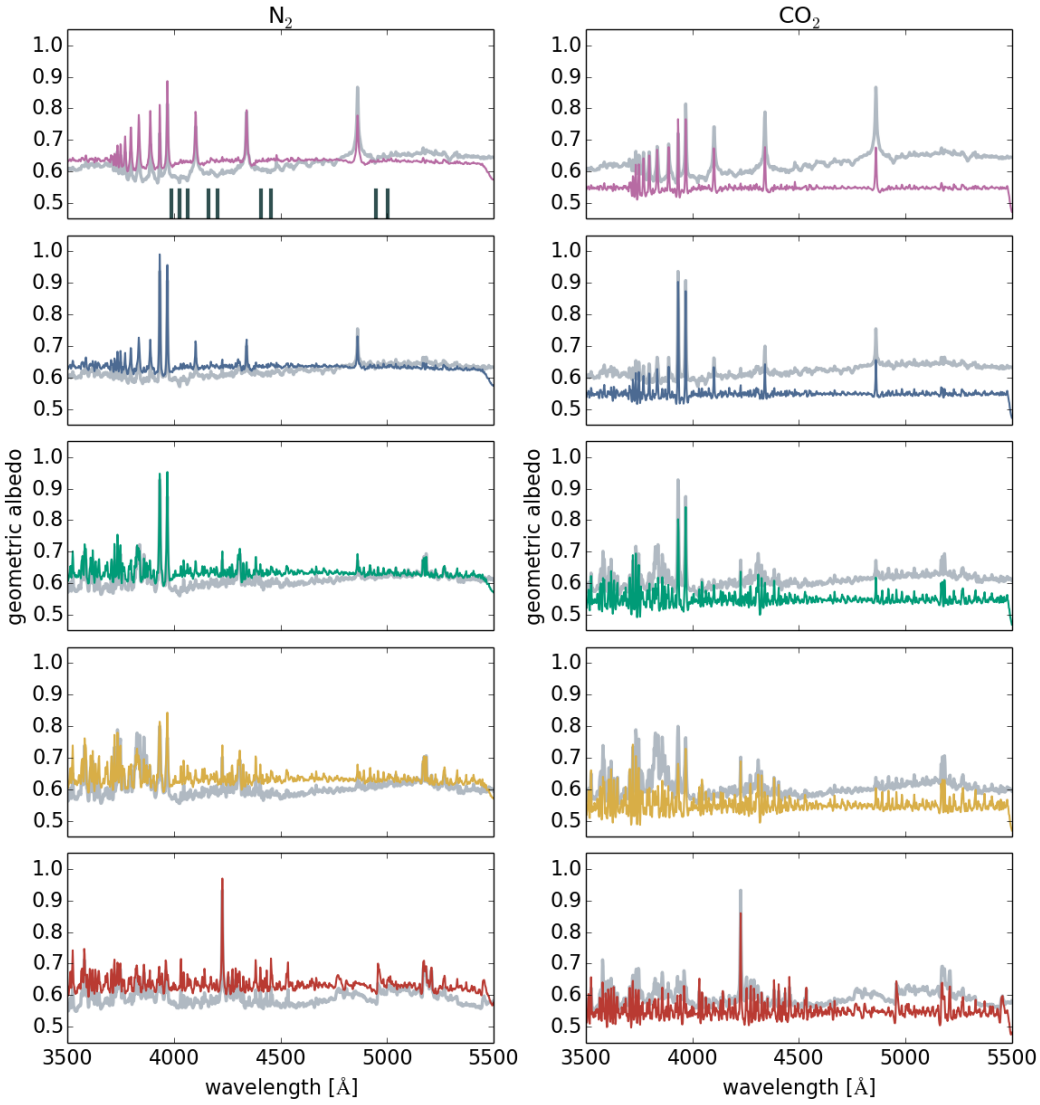

Figure 6.4: Geometric albedo spectra calculated for clear, 100 bar deep atmospheres of pure N_2 (left-hand side) and CO_2 (right-hand side), irradiated by stellar spectra shown in Figure 6.3. For comparison, the underlying gray lines show the corresponding albedo spectra from Figure 6.3, calculated for H_2/He composition. Dark bars in the top-left panel mark the central wavelengths of some of the most prominent hydrogen Raman ghost lines, corresponding to broad dips in the albedo spectra of hydrogen, which do not appear in the albedo spectra of heavy molecules at the same wavelengths.

6.4 Results and Discussion

6.4.1 Geometric albedo spectra

The geometric albedo spectra of the H_2/He model atmosphere, calculated for each of the five stellar spectra, are shown on the right-hand side of Figure 6.3. Note that Rayleigh scattering alone produces a smooth albedo spectrum, and thus all the features seen in the albedo spectra are the result of Raman scattering. The atmospheric model is identical in all five cases shown in Figure 6.3 and all variations in the albedo spectra are due to differences in the stellar spectra. The overall shape of the stellar spectrum and the strength of individual stellar lines change from one spectral type to another, and they both affect the intensity of Raman features.

Albedo peaks are caused by photons that have been Raman-scattered into the wavelengths of absorption lines, mostly from shorter wavelengths. The presence of absorption lines in the stellar spectrum is therefore a necessary condition for Raman albedo peaks to arise. The most prominent albedo peaks appear at wavelengths corresponding to the strongest lines in the stellar spectrum, which vary across different spectral types. Raman features at short wavelengths are more pronounced in an albedo spectrum of an atmosphere irradiated by a hot star, whose blackbody flux peaks at shorter wavelengths, compared to an atmosphere around a cool star, which is much fainter at blue wavelengths.

In the case of the earliest stellar type we consider, A7V, the most prominent features in the geometric albedo spectrum are associated with hydrogen Balmer lines: Hβ at 4861 Å, Hγ at 4340 Å, and Hδ at 4102 Å. In F-type stars, the Balmer lines become less pronounced, and the H and K lines of ionized calcium, at 3968 Å and 3933 Å, respectively, become the dominant features. For G-type stars, the albedo spectrum is dominated by the Ca II H and K lines[3], and the Raman peaks corresponding to the Balmer lines become even fainter than in the case of F stars. The magnesium feature at 5167 Å and the Ca I line at 4226 Å start to arise. In early K stars, the Ca II lines are still quite strong and the features due to magnesium and calcium appear more prominently. They become the dominant features in albedo spectra calculated for late K-types.

[3]The peaks in the albedo spectrum for the G5V star, shown in Figure 6.3, appear slightly less intense than the peaks produced by the solar spectrum in a similar planetary atmosphere shown in Chapter 5. This is mainly the result of different spectral resolution—here we compute albedo spectra on a wavelength grid with spacing twice as large compared to what was used in Chapter 5 (2 Å versus 1 Å). Raman features are created by individual stellar lines, and thus their sharpness depends on the spectral resolution.

In Figure 6.4, we show the geometric albedo spectra calculated for two additional atmospheric models, the atmospheres of pure N_2 (shown on the left-hand side) and pure CO_2 (on the right-hand side), irradiated by our five stellar spectra. The corresponding albedo spectra of the H_2/He atmosphere are shown for comparison in each panel (in gray). To first order, the albedo spectra of N_2 and CO_2 atmospheres look very similar to those calculated for the H_2/He model, having a continuum value of ~0.6 overlaid with Raman features at the wavelengths corresponding to prominent stellar lines. However, there are some notable differences arising from the fact that different molecules have different Rayleigh and Raman cross sections, as well as different values of Raman shifts.

Raman shift is the frequency offset imparted to the Raman-scattered photons. It corresponds to the energy difference between the initial and the final state of the molecule. For pure rotational Raman transitions, the initial and the final state belong to the same vibrational level of the molecule ($\Delta v = 0$), but differ in the rotational quantum number J by two ($\Delta J = \pm 2$). For ro-vibrational transitions, the initial and the final vibrational levels of the molecule are different ($\Delta v \neq 0$), and the rotational quantum number can either change by two or remain the same ($\Delta J = 0, \pm 2$). The energy difference between different rotational states of a molecule is inversely proportional to its moment of inertia; hence, the energy states in heavy molecules are more closely spaced than in light molecules. For that reason, Raman shifts are much smaller for heavy molecules, such as N_2 and CO_2 than for H_2, and the photons get re-distributed in frequency to a smaller degree. This explains why individual albedo peaks can have varying intensity depending on the scattering molecule, even though they generally seem to be of comparable size in all three types of atmospheres. Raman ghosts, on the other hand, are much more pronounced in the H_2/He model (the wavelengths of some of the most pronounced ghost lines are indicated by gray bars in the top-left panel in Figure 6.4) than in the other two cases, in which they are difficult to discern by eye because the ghosts and the albedo peaks are so close in wavelength that they almost overlap at this spectral resolution.

6.4.2 The effects of atmospheric absorption

The results shown in Figure 6.3 and Figure 6.4 are computed for deep, clear atmospheres (i.e. atmospheres with large gas column density) of very simplified composition without any absorbers. These are the conditions in which the effects of Raman scattering are strongest, hence they produce the most optimistic predictions for the intensity of Raman features in the albedo spectra of exoplanets.

Real atmospheres are expected to contain atomic and molecular absorbers, as well as clouds and hazes. The intensity of Raman features under these conditions is going to be reduced compared to our idealized case. In Chapter 5, we explored how the strength of Raman albedo features changes with varying the radiation-penetration depth of the atmosphere (related to the effective altitude of opaque clouds), the reflective properties of the could deck, the surface gravity of the planet, and the temperature-pressure profile of the atmosphere. Here we expand on our previous work and investigate the effect of absorption by atomic species present in the atmosphere.

The main reason why the geometric albedo is expected to be relatively high at short wavelengths is that this wavelength range does not contain many absorption lines/bands of various atmospheric constituents; instead, scattering is the main source of opacity. At longer visible and infrared wavelengths, the atmospheric opacity is dominated by absorbers such as water or methane, and the expected geometric albedo is therefore small.

However, there are a few notable absorption lines at short wavelengths associated with alkali metals sodium (Na) and potassium (K), which may affect the albedo spectrum of an exoplanet to a varying degree. The strength of the albedo absorption lines depends on the abundance of these atoms in monatomic form, as well as the ambient temperature and pressure. The mixing fractions of Na and K in giant planet atmospheres at high temperatures ($\gtrsim$ 1000 K) are close to solar values, $\sim 10^{-6}$ for Na, and $\sim 10^{-7}$ for K (see Figure 19 from Sharp & Burrows 2007). Below $\sim$ 1000 K the abundances start to rapidly decline due to the formation of condensates, which lock-in these elements and remove them from the gas phase.

To demonstrate how the geometric albedo spectrum of an exoplanet might be affected by alkali absorption, we perform a new set of radiative transfer calculations using modified versions of the H_2/He atmospheric model, irradiated by an early K-type stellar spectrum. We include Na and K absorption as additional sources of opacity (along with Rayleigh scattering on H_2 and He, and Raman scattering on H_2) in our calculation of the single-scattering albedo at each wavelength, which is used as input for DISORT radiative transfer code. The calculation of absorption cross sections for Na and K is described in Appendix 6.B. Figure 6.5 shows the albedo spectra of H_2/He-dominated atmospheres with sodium and potassium mixing ratios in the range $10^{-11} - 10^{-7}$. The bottom boundary of the model atmosphere shown in red is at a depth of 100 bar. In atmospheres with lower fractions of alkali metals (presumably

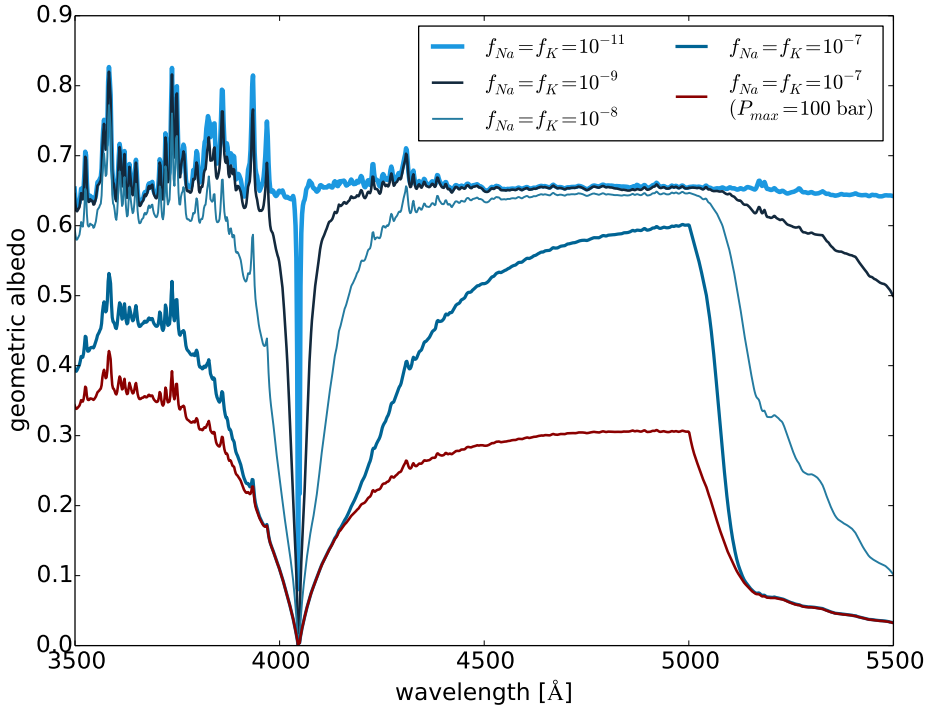

Figure 6.5: Geometric albedo spectra of atmospheres containing different mixing fractions of Na and K atoms. The atmospheres are irradiated by the spectrum of an early K star.

due to the formation of condensates), the bottom boundary is placed 'by hand' at the pressure level of 1 bar and it is assumed to be a partially absorbing surface with a wavelength-independent Lambertian albedo of 0.9. The bottom boundary properties are not calculated for any specific cloud model, they were selected to broadly represent two types of atmospheres—clear and cloudy—and to demonstrate that both the depth of the atmosphere and the mixing ratio of absorbers affect the strength of absorption lines which may preclude the formation of Raman spectral features.

The geometric albedos in Figure 6.5 show the effects of absorption lines of sodium at 3303 Å (which is not explicitly shown in the figure, but nevertheless it has an effect on the Raman-scattered light at longer wavelengths), of potassium at 4045 Å, and the Na-D resonance at 5890 Å, whose impact extends far away from the line center. For Na and K mixing fractions below $\sim 10^{-9}$, corresponding to atmospheric temperatures $\lesssim$ 800 K (Sharp & Burrows 2007), the absorption lines cause only a small change in the albedo spectrum compared to our idealized model

without any absorption (Figure 6.3). Therefore, alkali absorption can be safely ignored in moderately warm and cool atmospheres; however, it can inhibit the formation of prominent Raman spectral features in exoplanet atmospheres of very high temperature, such as those of hot Jupiters.

6.4.3 Observing the Raman features

The geometric albedo spectra at short optical wavelengths could, in principle, be measured using a few different observational techniques: (1) secondary eclipse observations, similar to the *STIS/HST* albedo measurements for HD 189733 b by Evans et al. (2013); (2) high-resolution spectroscopy with large ground-based observatories, similar to the technique used by Martins et al. (2015); and (3) direct spectroscopy using telescopes capable of achieving high contrasts at short optical wavelengths. There are no existing instruments in that category at the moment; however, a couple of concept studies for space-based missions equipped with coronagraphs and/or a starshades are under consideration, such as the *Habitable Exoplanet Imaging Mission (HabEx)* and the *Large UV/Optical/Infrared Surveyor (LUVOIR)*.

In Figure 6.6 we show the predicted signal-to-noise ratio (S/N) for direct detection of Raman features in the albedo spectrum of a nearby planet orbiting an early K star, HD 192310 c (Pepe et al. 2011). The properties of the planet-star system are listed in Table 6.2. The low atmospheric temperature expected for this planet justifies the use of the albedo spectrum computed without any absorption from Na and K. S/N is calculated using the instrument noise model for a space-based coronagraph described by Robinson et al. (2016), with some modifications: telescope diameter of 5 m (which is in the range of apertures proposed for *HabEx*), spectral resolution of $R = 200$, and a coronagraph capable of reaching the contrast level of 10^{-10}. We estimate S/N at which Raman features in the geometric albedo spectrum could be detected and distinguished from a smooth and featureless albedo spectrum of an atmosphere without any signatures of Raman scattering. We define 'signal' as the difference between the photon count coming from a planet with an albedo spectrum with prominent Raman features (solid line in the upper panel of Figure 6.6) and the photon count coming from a planet with a featureless albedo spectrum (dashed line). The predicted S/N distributions are computed for observations near quadrature (with the phase function value of $\phi = 0.5$), and for total integration time of 100 hr. Individual Raman features can be distinguished from a flat albedo spectrum with S/N of 2-3. Observing the spectral signatures of Raman scattering in the atmospheres of nearby exoplanets, such as HD 192310 c, is going to be an ambitious, but feasible

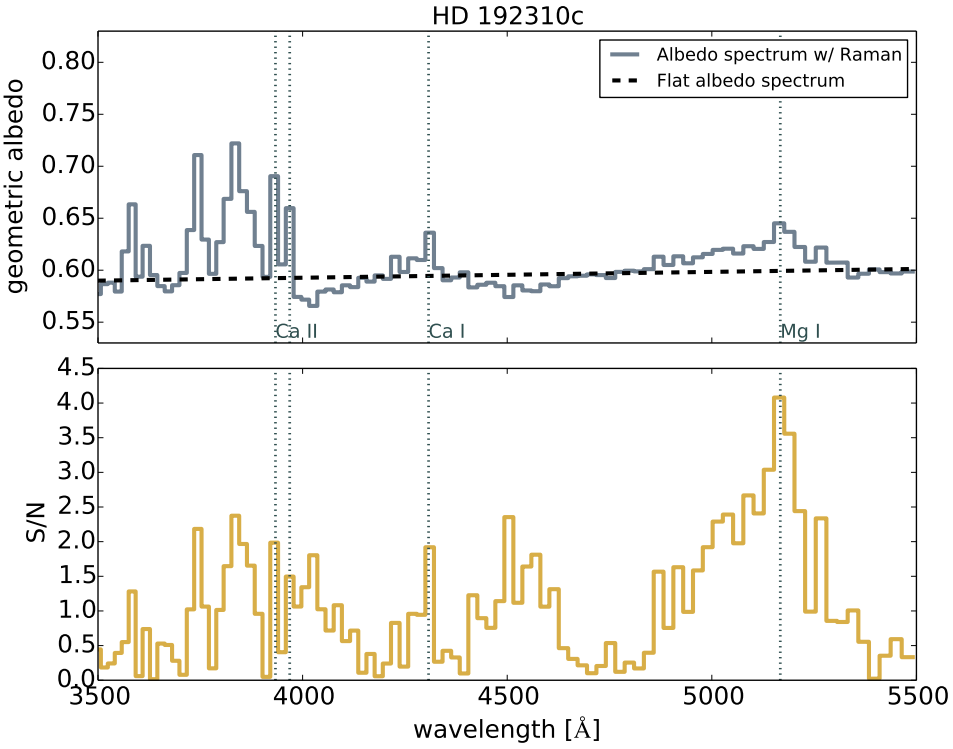

Figure 6.6: Upper panel: the solid gray line shows the predicted geometric albedo spectrum for an H_2/He atmosphere on a planet with properties of HD 192310 c, observed at $R = 200$ spectral resolution. The dashed line represents a smooth albedo spectrum without any Raman features. Observational signal is defined as the difference between these two curves. Lower panel: predicted S/N distribution as a function of wavelength for observations with a 5 m space telescope, for a total integration time of 100 hr, based on a modified version of the coronagraph noise model from Robinson et al. (2016).

task for the telescopes of next generation.

The prospects for observing Raman features in the albedo spectrum of a planet depend on both the strength of the features and on the brightness of the host star and the planet in the part of the spectrum where the features appear. Figure 6.6 also illustrates that in some case, especially in late-type stars, moderate-size Raman features (such as the Mg I albedo peak) can result in greater S/N than the most prominent albedo peaks, if the latter appear in a wavelength range in which the star is intrinsically faint. For G stars and earlier stellar types, the regions of highest S/N generally coincide with the most prominent Raman peaks.

Table 6.2: Properties of HD 192310 c (GJ 785 c) and its host star. References: (1) Pepe et al. (2011), (2) Nayak et al. (2017), (3) Fortney et al. (2007), (4) Gray et al. (2006), (5) Høg et al. (2000), (6) Howard et al. (2012).

Parameter	Value	Reference
$M_{pl} \sin i$	0.076 M_{Jup}	(1)
R_{pl}	0.75 R_{Jup}	(2, 3)
a	1.18 au	(1)
T_{eff}	185 K	(1)
spectral type	K2+V	(4)
V magnitude	5.723	(5)
distance	8.9 pc	(6)

6.5 Conclusions

Observations of the reflected light and the albedo spectra of exoplanets at short optical wavelengths ($3000 \lesssim \lambda \lesssim 6000$ Å) can provide insight into atmospheric properties. Contrary to a common assumption, the albedo at these wavelengths is not solely determined by Rayleigh scattering. Observations at moderate spectral resolution (R= $10^2 - 10^3$) can detect the spectral signatures of Raman scattering, which carry information about the depth of the atmosphere, its temperature, and composition.

Raman features do not only depend on the properties of the planetary atmosphere—they also depend on the shape of the stellar spectrum irradiating the planet. In this paper, we explore how the positions (i.e. wavelengths) and the intensities of different Raman features change due to the differences in the incident stellar light across different spectral types of host stars. For each stellar type, ranging from late A to late K stars, we identify the most promising Raman features that could be observed with the next generation of telescopes.

The recommended strategy for detecting and measuring Raman features in the albedo spectra of exoplanets would be to obtain moderate-resolution ($R = 10^2 - 10^3$) spectra over the wavelength range containing strong stellar lines, such as the Balmer lines or the Ca II H and K lines. The most suitable stellar types are the early types, which show strong Raman peaks and are also bright at short wavelengths. Late stellar types (such as our K7V case or later), although more numerous at close distance from the Sun, are intrinsically very faint at the wavelengths at which the most notable Raman peaks appear, thus making the observations of Raman features very challenging.

APPENDIX

6.A Rayleigh and Raman Scattering Cross Sections for CO_2

In order to compute CO_2 cross sections for Rayleigh scattering and the rotational Raman scattering, we use the approximate formulae from Dalgarno & Williams (1962). Because CO_2 has a relatively complicated structure of vibrational transitions, we consider only rotational Raman transitions in this work and assume all molecules are in the ground electronic and vibrational state.

For a transition between the initial molecular state of CO_2, characterized by the rotational quantum number J_i, to the final state, characterized by J_f, the cross section is

- for Rayleigh scattering ($J_i = J_f$):

$$Q(J_i; J_f) = \frac{128\pi^5}{9\lambda^4}\left[3\alpha^2 + \frac{2}{3}\frac{J_i(J_i+1)}{(2J_i-1)(2J_i+3)}\gamma^2\right] , \tag{6.2}$$

- for rotational S-branch transitions ($J_f = J_i + 2$):

$$Q(J_i; J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\frac{(J_i+1)(J_i+2)}{(2J_i+3)(2J_i+1)}\gamma^2 , \tag{6.3}$$

- for rotational O-branch transitions ($J_f = J_i - 2$):

$$Q(J_i; J_f) = \frac{128\pi^5}{9\lambda'\lambda^3}\frac{J_i(J_i-1)}{(2J_i-1)(2J_i+1)}\gamma^2 , \tag{6.4}$$

where λ and λ' are the outgoing and the incident wavelength, respectively. The incident and outgoing wavelengths are related by $\lambda'^{-1} = \lambda^{-1} + \Delta\nu$, where $\Delta\nu$ is the Raman shift of the transition in question, which is related to the energy difference between the initial and final state of the molecule involved in the scattering process. α and γ are the polarizability and the polarizability anisotropy of CO_2 at the equilibrium separation of nuclei, which have values of $\alpha = 2.3353 \times 10^{-24}$ cm^3 and $\gamma = 1.7915 \times 10^{-24}$ cm^3 (Morrison & Hay 1979). The differences between rotational energy levels are computed using the values of rotational constants for CO_2, $B = 0.3902$ cm^{-1} and $D = 0.12 \times 10^{-6}$ cm^{-1}, from Herzberg & Herzberg (1953).

6.B Absorption Cross Sections for Sodium and Potassium

The absorption cross section for an atomic transition from a lower state l to an upper state u is given by

$$\sigma_{lu}(\nu) = \frac{\pi e^2}{m_e c} f_{lu} \phi_\nu \,, \tag{6.5}$$

where e is the elementary electric charge, m_e is the electron mass, c is the speed of light, f_{lu} is the oscillator strength of the transition, and ϕ_ν is the line profile, normalized so that $\int \phi_\nu d\nu = 1$.

We adopt the Voigt line profile, which is a convolution of a Gaussian and a Lorentzian profile, computed using the PYTHON code described in Hill (2016)[4]. The line profile is calculated using the real part of the Faddeeva function $\omega(z)$ as

$$V(\nu - \nu_0; \sigma, \gamma) = \frac{\mathrm{Re}[\omega(z)]}{\sigma\sqrt{2\pi}} \,, \tag{6.6}$$

where $z = \frac{\nu - \nu_0 + i\gamma}{\sigma\sqrt{(2)}}$, ν_0 is the line center frequency, σ is the standard deviation of the Gaussian profile, and γ is the half-width at half-maximum (HWHM) of the Lorentzian.

The Gaussian part of the line profile originates from thermal Doppler broadening. Its HWHM is given by

$$\alpha = \nu_0 \sqrt{\frac{2\ln(2) k_B T}{mc^2}} \,. \tag{6.7}$$

It is related to the Gaussian standard deviation σ through $\alpha = \sigma\sqrt{2\ln 2}$.

The dominant part of the Lorentzain line profile is due to pressure broadening. We follow the approach described in Schweitzer et al. (1996) and model the pressure broadening of atomic lines in terms of the van der Waals (vdW) interaction between neutral, unpolarized particles. This interaction leads to a Lorentzian line profile with a HWHM of

$$\gamma = \frac{17}{2} C_6^{2/5} v_{\mathrm{rel}}^{3/5} n_{\mathrm{per}} \,, \tag{6.8}$$

where C_6 is the vdW interaction constant, v_{rel} is the relative velocity between the absorbing atom (Na or K) and the perturber, and n_{per} is the number density of the perturber. For v_{rel}, we use the expression for the mean relative speed between particles described by the Maxwell-Boltzmann distribution

$$\langle v_{\mathrm{rel}} \rangle = \sqrt{\frac{8 k_B T}{\pi \mu}} \,, \tag{6.9}$$

[4] http://scipython.com/book/chapter-8-scipy/examples/the-voigt-profile/

Table 6.3: Spectral line parameters.

Atom	Wavelength [nm]	f_{ik}	E_i [eV]	E_k [eV]
Na I	330.3320	0.009	0.0	3.753322
Na I	330.3930	0.00446	0.0	3.752628
K I	404.5279	0.00569	0.0	3.0649066
K I	404.8351	0.00263	0.0	3.062581

where $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass of the two particles.

We calculate the interaction constant as

$$C_6 = 1.01 \times 10^{-32}(Z+1)^2 \frac{\alpha_p}{\alpha_H}\left[\frac{E_H^2}{(E-E_l)^2} - \frac{E_H^2}{(E-E_u)^2}\right] \mathrm{cm}^6\mathrm{s}^{-1}\ , \tag{6.10}$$

where Z is the electric charge of the absorber, which we assume to be zero, since Na and K are mostly in their neutral states at the levels of the atmosphere that we are considering. α_p and α_H are the polarizabilities of the pertruber and the hydrogen atom, respectively. Since our model atmosphere is mostly made of H_2, we assume that this molecule is the dominant perturber and that all pressure broadening is due to interactions of the absorbing species with H_2. The polarizabilities relevant for our study are:

$$\alpha_H = 0.666793 \times 10^{-24}\ \mathrm{cm}^3, \tag{6.11}$$

$$\alpha_{H_2} = 0.806 \times 10^{-24}\ \mathrm{cm}^3. \tag{6.12}$$

E and E_H denote the ionization potential of the absorbing species and of the hydrogen atom, respectively:

$$E_\mathrm{H} = 13.6\ \mathrm{eV}, \tag{6.13}$$

$$E_\mathrm{Na} = 5.139\ \mathrm{eV}, \tag{6.14}$$

$$E_\mathrm{K} = 4.341\ \mathrm{eV}. \tag{6.15}$$

E_l and E_u are the excitation energies of the lower and upper level involved in the transition. There are four strong transitions (two doublets) in the wavelength range considered in our analysis, listed in Table 6.B. All atomic parameters are taken from the *NIST* Atomic Spectra Database[5].

Strong resonances of sodium and potassium at 5890 and 7700 Å can be important sources of opacity in exoplanet atmospheres at wavelengths far away from the line

[5] `http://physics.nist.gov/PhysRefData/ASD/lines_form.html`

centers. Even though we do not extend our analyzed wavelength range beyond 5500 Å, we include the absorption wings of the sodium-D resonance, which can affect the albedo spectra at wavelengths $\gtrsim$ 5000 Å. Describing the wings of these resonance lines with Lorentzian profiles has been shown to be inadequate. The Lorentz theory relies on the validity of the impact approximation (i.e. assuming instantaneous collisions between atoms and perturbers), which is not satisfied far from the line center of strong resonance lines. More rigorous treatments of the atom-perturber interaction produce line profiles with sharper cutoffs, stronger blue wings with satellite features, and a pronounced red/blue asymmetry (Burrows & Volobuyev 2003; Allard et al. 2012). We use the Na-D line profile from Allard et al. (2012), calculated for an atmospheric density of $n(H_2) = 10^{19}$ cm^{-3} and temperature 1500 K, made available by the authors in tabular form. Since the line profile is provided in terms of an absorption coefficient in arbitrary units, we re-scaled it so that the values in the blue wing match the cross section profile (in cm^2) from Burrows & Volobuyev (2003, Figure 6). The validity of the impact approximation (and consequently, the use of the Lorentzian profile in the line wings) for weaker sodium and potassium lines, such as those listed in Table 6.B, can also be called into question. This could be relevant in atmospheres of high mixing ratios of Na and K, which produce relatively broad absorption features. Calculating more accurate cross section profiles is beyond the scope of this paper; therefore we use the vdW Lorentzian profiles, but note that this may not be completely justified.

References

Allard, N. F., Kielkopf, J. F., Spiegelman, F., Tinetti, G., & Beaulieu, J. P., 2012, A&A, 543, A159, A159

Barstow, J. K., Aigrain, S., Irwin, P. G. J., et al., 2014, ApJ, 786, 154, 154

Belton, M. J. S., McElroy, M. B., & Price, M. J., 1971, ApJ, 164, 191

Belton, M. J. S., Wallace, L., & Price, M. J., 1973, ApJL, 184, L143

Bétremieux, Y., & Yelle, R. V., 1999, Icarus, 142, 324

Birkby, J. L., de Kok, R. J., Brogi, M., Schwarz, H., & Snellen, I. A. G., 2017, AJ, 153, 138, 138

Burrows, A., & Volobuyev, M., 2003, ApJ, 583, 985

Burrows, A., Ibgui, L., & Hubeny, I., 2008, ApJ, 682, 1277-1282, 1277

Cahoy, K. L., Marley, M. S., & Fortney, J. J., 2010, ApJ, 724, 189

Christiansen, J. L., Ballard, S., Charbonneau, D., et al., 2010, ApJ, 710, 97

Cochran, W. D., 1981, Advances in Space Research, 1, 143

Cochran, W. D., & Trafton, L. M., 1978, ApJ, 219, 756

Cochran, W. D., Trafton, L., Macy, W., & Woodman, J. H., 1981, ApJ, 247, 734

Courtin, R., 1999, Planet. Space Sci., 47, 1077

Cowan, N. B., & Agol, E., 2011, ApJ, 729, 54, 54

Dalgarno, A., & Williams, D. A., 1962, MNRAS, 124, 313

Demory, B.-O., 2014, ApJL, 789, L20, L20

Demory, B.-O., Seager, S., Madhusudhan, N., et al., 2011, ApJL, 735, L12, L12

Demory, B.-O., de Wit, J., Lewis, N., et al., 2013, ApJL, 776, L25, L25

Désert, J.-M., Charbonneau, D., Demory, B.-O., et al., 2011, ApJS, 197, 14, 14

Evans, T. M., Pont, F., Sing, D. K., et al., 2013, ApJL, 772, L16, L16

Fortney, J. J., Marley, M. S., & Barnes, J. W., 2007, ApJ, 659, 1661

Gray, R. O., Corbally, C. J., Garrison, R. F., et al., 2006, AJ, 132, 161

Heng, K., & Demory, B.-O., 2013, ApJ, 777, 100, 100

Heng, K., Mendonça, J. M., & Lee, J.-M., 2014, ApJS, 215, 4, 4

Herzberg, G., & Herzberg, L., 1953, Journal of the Optical Society of America (1917-1983), 43, 1037

Hill, C. 2016, Learning Scientific Programming with Python

Høg, E., Fabricius, C., Makarov, V. V., et al., 2000, A&A, 355, L27

Horak, H. G., 1950, ApJ, 112, 445

Horak, H. G., & Little, S. J., 1965, ApJS, 11, 373

Howard, A. W., Marcy, G. W., Bryson, S. T., et al., 2012, ApJS, 201, 15, 15

Hu, R., Ehlmann, B. L., & Seager, S., 2012, ApJ, 752, 7, 7

Karkoschka, E., 1994, Icarus, 111, 174

—, 1998, Icarus, 133, 134

Kipping, D., & Bakos, G., 2011, ApJ, 730, 50, 50

Latham, D. W., Borucki, W. J., Koch, D. G., et al., 2010, ApJL, 713, L140-L144, L140

Madhusudhan, N., & Burrows, A., 2012, ApJ, 747, 25, 25

Marley, M. S., Gelino, C., Stephens, D., Lunine, J. I., & Freedman, R., 1999, ApJ, 513, 879

Martins, J. H. C., Santos, N. C., Figueira, P., et al., 2015, A&A, 576, A134, A134

Morley, C. V., Fortney, J. J., Marley, M. S., et al., 2015, ApJ, 815, 110, 110

Morrison, M. A., & Hay, P. J., 1979, JChPh., 70, 4034

Nayak, M., Lupu, R., Marley, M. S., et al., 2017, PASP, 129.3, 034401

Oklopčić, A., Hirata, C. M., & Heng, K., 2016, ApJ, 832, 30

Pepe, F., Lovis, C., Ségransan, D., et al., 2011, A&A, 534, A58, A58

Pickles, A. J., 1998, PASP, 110, 863

Prather, M. J., 1974, ApJ, 192, 787

Price, M. J., 1977, Reviews of Geophysics and Space Physics, 15, 227

Robinson, T. D., Stapelfeldt, K. R., & Marley, M. S., 2016, PASP, 128.2, 025003

Rowe, J. F., Matthews, J. M., Seager, S., et al., 2008, ApJ, 689, 1345-1353, 1345

Schweitzer, A., Hauschildt, P. H., Allard, F., & Basri, G., 1996, MNRAS, 283, 821

Sharp, C. M., & Burrows, A., 2007, ApJS, 168, 140

Snellen, I. A. G., de Mooij, E. J. W., & Albrecht, S., 2009, Nature, 459, 543

Sromovsky, L. A., 2005a, Icarus, 173, 284

—, 2005b, Icarus, 173, 254

Stamnes, K., Tsay, S.-C., Jayaweera, K., & Wiscombe, W., 1988, ApOpt, 27, 2502

Sudarsky, D., Burrows, A., & Hubeny, I., 2003, ApJ, 588, 1121

Valdes, F., Gupta, R., Rose, J. A., Singh, H. P., & Bell, D. J., 2004, ApJS, 152, 251

Wallace, L., 1972, ApJ, 176, 249

Yelle, R. V., Doose, L. R., Tomasko, M. G., & Strobel, D. F., 1987, Geophys. Res. Lett., 14, 483